

Scientific and Technical Information Center

SEARCH REQUEST FORM

Date: _____ Requester's Full Name: _____ Examiner #: _____
Art Unit: _____ Phone (30____) _____ Serial Number: _____
Results Format Preferred (circle): PAPER DISK E-MAIL

To ensure an efficient and quality search, please attach a copy of the cover sheet, claims, and abstract or fill out the following:

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Date: _____

Search Topic:

Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known.

For Sequence Searches Only Please include all pertinent information (parent, grandchild, divisional, or issued patent numbers) along with the appropriate serial number.

STAFF USE ONLY

Searcher: M. S. M. 83278

Searcher Phone #: _____

Searcher Location: _____

Date Searcher Picked Up: _____

Date Completed: 11-24-99

Searcher Prep & Review Time: 10

Online Time: 15

Type of Search

____ NA Sequence (#)

1 AA Sequence (#)

____ Structure (#)

____ Bibliographic

____ Litigation

____ Fulltext

____ Other

Vendors and Cost

____ STN _____ Dialog

____ Questel/Orbit _____ Dr. Link

____ Lexis/Nexis _____ Westlaw

____ WWW/Internet

____ In-house sequence systems (list)

____ Other (specify)

CompuGen

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 1999, 20:28:01 ; Search time 15.28 Seconds
(without alignments)
942.484 Million cell updates/sec

Title: US-09-095-385-4
 Perfect score: 3226
 Sequence: 1 MLLFVLTCLLAVFPAISTKS.....SGFREIENKAIQDPRFLFAEZ 608

Scoring table: BLOSUM62

Searched: 188963 seqs, 23686106 residues

Database : A_Geneseq_36:★

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query			ID	Description	
	Score	Match	Length			
1	322.9	99.9	607	1	W95601	Human secretory Im
2	3139	97.3	746	1	W03178	Human poly-immunog
3	2177	67.5	757	1	W03179	Bovine poly-immunog
4	2099.5	65.1	771	1	W03180	Mouse poly-immunog
5	2064.5	64.0	769	1	W03181	Rat poly-immunoglo
6	1435.5	44.5	773	1	W03177	Rabbit poly-immuno
7	1431	44.4	584	1	R14670	Truncated poly Ig-
8	1248	38.7	584	1	R14671	Truncated poly Ig-
9	331	10.3	60	1	W99073	Immunoglobulin lik
10	287.5	8.9	57	1	W99072	Immunoglobulin lik
11	261.5	8.1	532	1	W50033	Human immunity rel
12	232	7.2	43	1	W99078	Immunoglobulin lik
13	215	6.7	46	1	W99077	Immunoglobulin lik
14	207.5	6.4	592	1	W94269	H6FXscfV(CEA6)trp
15	195.5	6.1	1910	1	W42086	Human Down syndrom
16	194.5	6.0	1571	1	W48089	Human Down syndrom
17	184	5.7	483	1	W88099	A protein comprisi
18	183	5.7	486	1	R37649	Sequence of divale
19	182	5.6	390	1	Y05001	Human PIGRL-1 prot
20	173.5	5.4	553	1	R56966	CC49 VL-L-VH-L-VL-
21	173.5	5.4	553	1	W91780	A multivalent sing
22	172.5	5.3	553	1	W11508	Single chain anti-
23	172.5	5.3	533	1	W73223	H22-anti-CEA antib
24	172	5.3	580	1	W90217	Bispecific tetra
25	171.5	5.3	556	1	W90218	Bispecific tetra
26	169.5	5.3	647	1	R39549	Specific monoclon
27	168.5	5.2	739	1	R08118	Vascular cell adhe
28	168.5	5.2	739	1	R0316	VE7/2G7 staloglyco
29	168.5	5.2	847	1	R33419	CD22-beta. Inhibit
30	168.5	5.2	739	1	R38548	VCAM-7D. Monoclon
31	168.5	5.2	735	1	R38550	VCAM/ICAM-1. Monoc
32	168.5	5.2	736	1	R39682	VCAM/ICAM-2. Monoc
33	168.5	5.2	739	1	R49733	Sequence of the IE
34	168.5	5.2	739	1	W46734	Amino acid sequenc
35	168	5.2	739	1	R13906	Human VCAM-1. New
36	167.5	5.2	739	1	T21081	VCAM variant with
37	166.5	5.2	1020	1	W05141	scFv2(FRP5/225)-ET
38	164	5.1	1021	1	W32381	Alloreaction assoc
39	163.5	5.1	647	1	R08117	Vascular cell adhe
40	163.5	5.1	828	1	R13905	Rabbit ATHERO-ELAM
41	163.5	5.1	211	1	W86306	Kidney injury asso
42	163	5.1	1028	1	W29667	Homo sapiens DL185
43	160.5	5.0	739	1	R27456	1E7-267 antigen. R

ALIGNMENTS

RESULT	1
--------	---

W95601
ID W95601 standard; Protein; 607 AA.

AC	W95601;	
AD	08-JUN-1999	(first entry)
DE	Human secretory Immunoglobulin A component.	
DE	Immunoglobulin A; secretory; component; IgA; human; treatment;	
KW	prevention; infection; HIV; AIDS; cold; flu; virus;	
KW	human immunodeficiency virus; respiratory syncytial virus.	
OS	Homo sapiens.	
PN	W09857993-Al.	
PD	23-DEC-1998.	
PF	10-JUN-1998;	U11975.
PR	19-JUN-1997;	DS-050969.
PA	(REGC) UNIV CALIFORNIA.	
PI	Chintalacharuvu KR, Morrison SL;	
PI	WPI; 99-080950/07.	
DR	N-PSDA: X07407.	

PT produce secretory immunoglobulin in single cells - useful to
PS produce commercial quantities of secretory immunoglobulin to prevent
or treat infections
PS Disclosure: Pages 22-24; 39pp; English.

The sequence is that of the secretory component of human secretory immunoglobulin A (sigA). It can be used as part of a method for the production of sig molecules. This method is useful for producing commercial quantities of sig (especially sigA) to treat or prevent infections. In particular, sigA produced by the method can be used to prevent or treat infections in mammals, birds or fish; especially systemic infections or infections at a mucosal surface. It is especially useful to prevent or treat infection with human immunodeficiency virus (HIV), respiratory syncytial virus, flu virus or cold virus. The method allows production of commercial quantities of sig molecules for therapeutic use, not previously possible; production using non-plant cells and a single cell type is more efficient than a previous multi-step process of fusing recombinant plant cells, and avoids alterations of the sig by plant cells. SigA molecules, are more stable and resistant to proteolysis than previously used IGA molecules, and can be administered to prevent as well as to treat infections, unlike e.g. IgG and IgM molecules.

Sequence	607 AA;
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	
101	
102	
103	
104	
105	
106	
107	
108	
109	
110	
111	
112	
113	
114	
115	
116	
117	
118	
119	
120	
121	
122	
123	
124	
125	
126	
127	
128	
129	
130	
131	
132	
133	
134	
135	
136	
137	
138	
139	
140	
141	
142	
143	
144	
145	
146	
147	
148	
149	
150	
151	
152	
153	
154	
155	
156	
157	
158	
159	
160	
161	
162	
163	
164	
165	
166	
167	
168	
169	
170	
171	
172	
173	
174	
175	
176	
177	
178	
179	
180	
181	
182	
183	
184	
185	
186	
187	
188	
189	
190	
191	
192	
193	
194	
195	
196	
197	
198	
199	
200	
201	
202	
203	
204	
205	
206	
207	
208	
209	
210	
211	
212	
213	
214	
215	
216	
217	
218	
219	
220	
221	
222	
223	
224	
225	
226	
227	
228	
229	
230	
231	
232	
233	
234	
235	
236	
237	
238	
239	
240	
241	
242	
243	
244	
245	

```
Query Match          99.9%; Score 3222; DB 1; Length 607;
Best Local Similarity 100.0%; Pred. No. 5.3e-240;
Matches 607; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

Qy	1	MLLFVTLCLLAVPAISTKSP	IFGPEVNSV	EGNSV	STCY	PPTSV	NRHTRK	KWC	RQGA	60			
Db	1	MLLFVTLCLLAVPAISTKSP	IFGPEVNSV	EGNSV	STCY	PPTSV	NRHTRK	KWC	RQGA	60			
Qy	61	RGCGITLISSEGYSSKY	AGRANLTN	FPENG	TGTFV	VNIAQL	SQDSD	SGRYK	CGLGIN	SGLS			
Db	61	RGCGITLISSEGYSSKY	AGRANLTN	FPENG	TGTFV	VNIAQL	SQDSD	SGRYK	CGLGIN	SGLS			
Qy	121	FDVSLVSV	QSGPGLLND	TKVYTV	DLGRTV	TINCP	FTENA	QAKR	KSLSY	KOIGLYPVLV	IDSS		
Db	121	FDVSLVSV	QSGPGLLND	TKVYTV	DLGRTV	TINCP	FTENA	QAKR	KSLSY	KOIGLYPVLV	IDSS		
Qy	181	GYPNPNT	GRTRID	IOGTG	QLLSV	VINQL	RLSD	AGQYL	COAGD	SDSN	KNKADQLV	LKP	
Db	181	GYPNPNT	GRTRID	IOGTG	QLLSV	VINQL	RLSD	AGQYL	COAGD	SDSN	KNKADQLV	LKP	
Qy	241	EPELYED	LRGSVTF	HCALG	PEV	ANYAK	ELCRQ	SSG	ENC	DVVYV	TLG	KRAPAF	EGRILLN
Db	241	EPELYED	LRGSVTF	HCALG	PEV	ANYAK	ELCRQ	SSG	ENC	DVVYV	TLG	KRAPAF	EGRILLN

```

QY 301 PODKGSFVWITGLRKEDAGRYLCAHSDGQLQEGSPIQAWQLFVNEESTIPRSPYVK 360
DB 301 PODKGSFVWITGLRKEDAGRYLCAHSDGQLQEGSPIQAWQLFVNEESTIPRSPYVK 360
QY 361 GVAGSSVAVLCPYNNRKSISIKYWLWEGAQNGRCPDLLVDSEGWYKAQYEGRLSLEEPG 420
DB 361 GVAGSSVAVLCPYNNRKSISIKYWLWEGAQNGRCPDLLVDSEGWYKAQYEGRLSLEEPG 420
QY 421 NGFTFVILNQLSRDAGFVWCLTNGDTLWRTTVEIKIIEGEPNLKVPNGNTAVLGETLKY 480
DB 421 NGFTFVILNQLSRDAGFVWCLTNGDTLWRTTVEIKIIEGEPNLKVPNGNTAVLGETLKY 480
QY 481 PCHFPCKFSSEYKYWKWNNTGQALPSQDEGSPKAFVNCDSNLSVLTNLVTRADEG 540
DB 481 PCHFPCKFSSEYKYWKWNNTGQALPSQDEGSPKAFVNCDSNLSVLTNLVTRADEG 540
QY 541 WYWCYKQGHFYGETAAVYVAVEERKAAGSRDVS LAKADAAPDEKVLDSGFREIENKAIQ 600
DB 541 WYWCYKQGHFYGETAAVYVAVEERKAAGSRDVS LAKADAAPDEKVLDSGFREIENKAIQ 600
QY 601 DPLRFAE 607
DB 601 DPLRFAE 607

RESULT 2
W03178
ID W03178 standard; Protein: 746 AA.
AC W03178;
DT 24-FEB-1997 (first entry)
DE Human poly-immunoglobulin receptor.
KW Human; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus.
OS Homo sapiens.
FH Key
FT region
FT /note= "putative immunoglobulin binding residues
FT of domain I"
FT 1..120
FT /label= domain_I
FT 110..230
FT /label= domain_II
FT 210..340
FT /label= domain_III
FT 320..450
FT /label= domain_IV
FT 440..550
FT /label= domain_V
FT 550..606
FT /note= "external portions of domain VI"
FT 550..627
FT /note= "external portions of domain VI"
FT 625..660
FT /label= transmembrane_segment
FT 650..746
FT /label= intracellular_portion
PN W09621012-A1.
PD 11-JUL-1996.
PR 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNNE-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PI (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI: 96-333987/33.
DR N-PSDB: T31288.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,

```

```

PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure; Pages 105-108; 152pp; English.
CC The present sequence is the human poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 746 AA;

Query Match 97.3%; Score 3139; DB 1; Length 746;
Best Local Similarity 99.8%; Pred. No. 1.7e-233;
Matches 589; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 19 KSPIFGPEEVNSVEGNSVITCYPPPTSVAHRHTRKYRCQARGGCITLISSEGYVSSKY 78
DB 1 KSPIFGPEEVNSVEGNSVITCYPPPTSVAHRHTRKYRCQARGGCITLISSEGYVSSKY 60
QY 79 AGRANLNFPENGTFVYVNIQALSDQDSGRYKCGLINSRGLSFDVSLVSGPGLLNDTK 138
DB 61 AGRANLNFPENGTFVYVNIQALSDQDSGRYKCGLINSRGLSFDVSLVSGPGLLNDTK 120
QY 139 VYTVDLGRTVTINCPFKTENAKRKSILYKQIGLYPVLVIDSSGYVNNYTRIRLDIOGT 198
DB 121 VYTVDLGRTVTINCPFKTENAKRKSILYKQIGLYPVLVIDSSGYVNNYTRIRLDIOGT 180
QY 199 GQLLFSVVIQLRLSDAGQYLQAGDDSNKKNADQLVKPEPELYEDLGRSVTFHCA 258
DB 181 GQLLFSVVIQLRLSDAGQYLQAGDDSNKKNADQLVKPEPELYEDLGRSVTFHCA 240
QY 259 LGPEVANVAKFLCROSSGECNDVVNTLGRAPAFEGRIILLNPODKGFSFVITGLRKE 318
DB 241 LGPEVANVAKFLCROSSGECNDVVNTLGRAPAFEGRIILLNPODKGFSFVITGLRKE 300
QY 319 DAGRYLCAHSDGQLQEGSPIQAWQLFVNEESTIPRSPYVKGVAGSSVAVLCPYNNRKS 378
DB 301 DAGRYLCAHSDGQLQEGSPIQAWQLFVNEESTIPRSPYVKGVAGSSVAVLCPYNNRKS 360
QY 379 KSIKYWLWEGAQNGRCPDLLVDSEGWYKAQYEGRLSLEEPGNGTFTVILNQLTSRDAGF 438
DB 361 KSIKYWLWEGAQNGRCPDLLVDSEGWYKAQYEGRLSLEEPGNGTFTVILNQLTSRDAGF 420
QY 439 YWCLTNGDTLWRTTVEIKIIEGEPNLKVPNGNTAVLGETLKVCPCHFPCKFSSEYKYWKW 498
DB 421 YWCLTNGDTLWRTTVEIKIIEGEPNLKVPNGNTAVLGETLKVCPCHFPCKFSSEYKYWKW 480
QY 499 NNTGCQALPSQDEGSPKAFVNCDSNLSVLTNLVTRADEGWYWCYKQGHFYGETAAV 558
DB 481 NNTGCQALPSQDEGSPKAFVNCDSNLSVLTNLVTRADEGWYWCYKQGHFYGETAAV 540
QY 559 YVAVEERKAAGSRDVS LAKADAAPDEKVLDSGFREIENKAIQDPLRFAE 608
DB 541 YVAVEERKAAGSRDVS LAKADAAPDEKVLDSGFREIENKAIQDPLRFAE 590

RESULT 3
W03179
ID W03179 standard; Protein: 757 AA.
AC W03179;
DT 24-FEB-1997 (first entry)
DE Bovine poly-immunoglobulin receptor.
KW Bovine; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;

```


PI Hiatt AC, Lehner T, Ma JKC;
DR WPI: 96-333987/33.
DR N-PSDB: T31290.
PT Immunoglobulin and protection protein complex and its prodn. in
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure: Pages 117-121; 152pp; English.
CC The present sequence is the mouse poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 771 AA;

Query Match 65.1%; Score 2099.5; DB 1; Length 771;
Best Local Similarity 64.2%; Pred. No. 1.8e-153;
Matches 395; Conservative 81; Mismatches 130; Indels 9; Gaps 5;

QY 1 MLLFVLTCLLAVFPAISTKPIGPEEVNSVEGNSVITCYIPTSVNRHTRYWCROGA 60
DB 1 MRLYLFTLLTVFSGVSTKPIFGPQEVSEIGSDSVITCYIPTSVNRHTRYWCROGA 60
QY 61 RGGCITLISSEGVSSKYAGRANLTNPENGTFVYVNIQALSDQSDSGRYKCGLGINSRGLS 120
DB 61 SGMCTTLLISNGYLSKEYSGRANLTNPENFTFVNIETQLTQDDTSGYKCGLGTSNRGLS 120
QY 121 FDSVLEVSQPGLLNDKVTYVDLGRVITNCPEKTENAKRSLYKQIGLYPVLVDS 180
DB 121 FDSVLEVSQVPELSDPTVHTYKIDGRNVTIECPFKRENVPFSKSLCKRTQSCVELVDST 180
QY 181 GYVNPNTGTRLDIOGTGQLLFVSVINOLRLSDAGOLQAGDSDSNKKNADLOVLKP 240
DB 181 EKNPSTIGRAKLFMKGTDLTFVYVNIHSHTHNDAGLYICQAGEGSPADKKNVDLQVLAP 240
QY 241 EPELYEDLRGVSFTFHCALGPEVANAKFLCRQSSGENDCVVNTLGRAPAFEGRIILN 300
DB 241 EPELYKDLRSSVTFECDLGREVANAKYLCRMNK-ETCDVINTLGRDPDPEGRILIT 299
QY 301 PQDKGFSVVITGLRKEDAGRYLCGRHSDGQLOEGSPIQAWQLFVNEESTIPRSPVVK 360
DB 300 PKDNGRFSVLITGLRKEDAGHYCQGAHSGSLPOEGWPQTQWLFVNEESTIPNRRSVK 359
QY 361 GVAGSSVAVILCPYNRKESKTKYCLWEGAGNCRCLLVDSEGVWKAQYEGRLSLLEPG 420
DB 360 GVTGGSVAIACPPYKPKSSSLKYCRWEGDNGHCPALVGTQAOVQEGYEGRLALFDQPG 419
QY 421 NGFTVILNQLTSDRAGFYWCLTNGDTLWRTTVEIKIEG--EPNLKV-PGNVTAVLGET 477
DB 420 NGFTVILNQLTSDRAGFYWCLTNGDSRWRITIELVAEATREPNEVLTQNTAVLGET 479
QY 478 LKVPCHFPCKFSSEYKWKWNKNTGQALPSQDDEGSPKAFVNCDENSRLVSLNLVTRA 537
DB 480 FTVSCHYPCKFYQEKYQWCKSNKSGCHILFSDHDEGARQSSVCDQSSQLVMTLNPYSKE 539
QY 538 DEGVYWCVGKQGHYFGTAAVYVAVERKAAGSRDV---SLAKADAAPDEKVLDSGFE 593
DB 540 DEGVYWCVGKQGHYFGTAAVYVAVERKAAGSRDV---SLAKADAAPDEKVLDSGFE 598
QY 594 IENKAIDPRLFAEZ 608
DB 599 KENKAIPNPGPFANE 613

RESULT 5

W03181
ID W03181 standard; Protein: 769 AA.
AC W03181:
DT 24-FEB-1997 (first entry)
DE Rat poly-immunoglobulin receptor.
KW Rat; immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus.
KW Rattus rattus.
FH Key
FT Location/Qualifiers
FT 13..45
FT /note= "putative immunoglobulin binding residues
of domain I"
FT domain
FT 1..120
FT /label= domain_I
FT domain
FT 110..230
FT /label= domain_II
FT domain
FT 210..340
FT /label= domain_III
FT domain
FT 320..450
FT /label= domain_IV
FT domain
FT 440..550
FT /label= domain_V
FT region
FT 550..606
FT /note= "external portions of domain VI"
FT region
FT 550..627
FT /note= "external portions of domain VI"
FT region
FT 625..660
FT /label= transmembrane_segment
FT region
FT 650..769
FT /label= intracellular_portion
W09621012-A1.
PN 11-JUL-1996.
PD 27-DEC-1995; U16889.
PR 30-DEC-1994; US-367395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (UNNE-) PLANT BIOTECHNOLOGY INC.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hiatt AC, Lehner T, Ma JKC;
DR WPI: 96-333987/33.
DR N-PSDB: T31291.
PT Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Disclosure: Pages 123-127; 152pp; English.
CC The present sequence is the rat poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, pref.
CC 1-606, or esp. residues 13-45, 1-120, 110-230, 210-340, 320-450,
CC 440-550, 550-606 or 550-627 comprises a protection protein (PP).
CC The Ig of the invention comprises a PP as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The PP protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
SQ Sequence 769 AA;

Query Match 64.0%; Score 2064.5; DB 1; Length 769;
Best Local Similarity 62.6%; Pred. No. 8.6e-151;
Matches 385; Conservative 96; Mismatches 125; Indels 9; Gaps 5;

QY 1 MLLFVLTCLLAVFPAISTKPIGPEEVNSVEGNSVITCYIPTSVNRHTRYWCROGA 60
DB 1 MRLSLFALLTVFSGVSTKPIFGPQDVSEIGNSVITCYIPTSVNRHTRYWCROGA 60
QY 61 RGGCITLISSEGVSSKYAGRANLTNPENGTFVYVNIQALSDQSDSGRYKCGLGINSRGLS 120

```

Db 61 NGYCATLISSNGYLSKEYSGRASLINFENSTFVNIAHLAQEDTGSYKCGLGTNRGLF 120
QY 121 FDSVLEYSOGPLGNDTKVYVVDLGRVTINCPFKTENAKRKSLYKQIGLYPVLYDSS 180
Db 121 FDSVLEYSQVPEFNDTHVYTKDGRVTIEICRFKEGNAHKSLSCKRKGACEVVIDST 180
QY 181 GYVNPNTGRILDIQGTGQLFSVYVINOIRLSDAGQYLCQAGDDSNKKNADLQVLKP 240
Db 181 EYVDPYSKDRALLFMKGTSDRFYVYNISHLIPSDAGLVVCQAGEGPSADKKNADLQVLEP 240
QY 241 EPELVYEDLRSVTFHCALGPEVANAFKFLCRQSGGNCVYVNTLGRAPAFGRILLN 300
Db 241 EPELVYEDLRSVTFHCALGPEVANAFKFLCRQSGGNCVYVNTLGRAPAFGRILLN 299
QY 301 PODKGSFVITGLRKEDAGRYLCGAHSDGLOEGSPIQAWOLFVNEESTIPRSPTVVK 360
Db 301 PRDDNGRFSLITGLRKEDAGHYCGAHSGLPQEGWPVQAWOLFVNEESTIPRSVVK 359
QY 361 GVAGSSVAVLCPYNRKESKIKYWCLEWGAQNGRCPLLDVDSGWWKAYQYGRLSLEEPG 420
Db 361 GVTGSSVAVLCPYNRKESKIKYWCLEWGAQNGRCPLLDVDSGWWKAYQYGRLSLEEPG 419
QY 421 NGTFTVILNQLTSRDAGFYWCLTNGDILWRTVTEIKIIEG--EPNLKY-PCNVTAVLGET 477
Db 420 SGAYTVILNQLTQDSGFYWCLTNGDILWRTVTEIKIIEG--EPNLKY-PCNVTAVLGET 479
QY 478 LKVPCHPCKFESSYKWKCKNNNTGCOALPSODEGSPKAFKNCNDSRLVSLTLNLVTRA 537
Db 480 FTISCHPCKFYSEKWKCKNNNTGCOALPSODEGSPKAFKNCNDSRLVSLTLNLVTRA 539
QY 538 DEGWYWCYKOGHYFGETAAYVAVEERKAAGSRDVSIAKADA----APDEKVLDSGFR 593
Db 540 DEGWYWCYKOGHYFGETAAYVAVEERKAAGSRDVSIAKADA----APDEKVLDSGFR 598
QY 594 IENKAIQDPRLFAEZ 608
Db 599 DENKANLDPRLFADE 613

RESULT 6
ID W03177 standard; Protein: 773 AA.
AC W03177;
DE Rabbit poly-immunoglobulin receptor.
KW Rabbit poly-immunoglobulin; receptor; protection protein; mutants;
KW heavy chain; antigen binding domain; protection; pathogen;
KW mucosal; environment; gastrointestinal; passive; immunisation;
KW Guy's 13 antibody; prevention; dental caries; Streptococcus;
KW poly; sorbinus.
OS Oryctolagus cuniculus.
FH Key Location/Qualifiers
FT region 21..43
FT /note="immunoglobulin binding residues of domain I"
FT domain 1..118
FT /label=domain_I
FT domain 119..223
FT /label=domain_II
FT domain 224..332
FT /label=domain_III
FT domain 333..441
FT /label=domain_IV
FT domain 442..552
FT /label=domain_V
FT region 553..606
FT /note="external portions of domain VI"
FT region 553..627
FT /note="external portions of domain VI"
FT region 630..652
FT /label=transmembrane_segment
FT region 653..755

```

```

FT W09621012-A1. /label= intracellular_portion
PN 11-JUL-1996.
PF 27-DEC-1995; U16889.
PR 30-DEC-1994; US-387395.
PR 04-MAY-1995; US-434000.
PA (PLAN-) PLANT BIOTECHNOLOGY INC.
PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
PA (PLAN-) PLANET BIOTECHNOLOGY INC.
PI Hlatt AC, Lehner T, Ma JKC;
PI WPL: 96-333987/33.
DR N-PSDB; T31287.
DR Immunoglobulin and protection protein complex and its prodn. in
PT plants - useful for passive immunisation against mucosal antigens,
PT esp. against S. mutans and S. sorbinus to prevent dental caries
PS Claim 10; Pages 99-102; 152pp; English.
CC The present sequence is the rabbit poly-immunoglobulin (Ig)
CC receptor, a portion of which corresp. to residues 1-627, Pref.
CC 1-606, or esp. residues 21-43, 1-118, 119-223, 224-332, 333-441,
CC 442-552, 553-606 or 553-627 comprises a pp as above in association with
CC the Ig of the invention comprises a pp as above in association with
CC an Ig derived heavy chain, having at least a portion of an antigen
CC (Ag) binding domain. The pp protects the Ig in harsh mucosal, e.g.
CC gastrointestinal, environments, therefore enhancing its
CC effectiveness in passively immunising animals against mucosal
CC pathogens. The Ag binding domain is specifically derived from the
CC Guy's 13 antibody, and the Ig can be used to prevent dental caries
CC by binding, e.g. Streptococcus mutans serotypes c, e and f, or
CC S. sorbinus serotypes d and g.
CC Sequence 773 AA;
QY 1 MLLFVLTCLLAVFPAISTK-----SPIFGPEVNSVEGNSVITCYPTSVNRHTRKY 54
Db 1 MALFLTCLLAVFSAATAQSSLLGPSIFGPEVNVLEGDSVITCYPTSVTRHSRKF 60
QY 55 WCROGARGGGITLISSEGYVSSKYAGRANLTNPENGTFFVNTIAQLSQDSDGRKCGLGI 114
Db 61 WCREESGRCVTL-ASTGYSQTSYSGRGLTDFPDKGFEVYVTDQLTQNDSGSYKCGVG 119
QY 115 NSRGLSDVLSVYSGQPLGNDTKVTVDLGRVTINCPFKTENAKRKSLYKQIGLYPV 174
Db 120 NGRGLDFGVNVLVSQKPE--PDDVYKQYESYVTITCPTYATRLKSKFYKVEDGELV 177
QY 175 LVIDSQ--YVNPNTGRIRLDIQGTGQLFSVYVINOIRLSDAGQYLCQAGDDSNKKN 232
Db 178 LIIDSSKKAQDPYKGRITLQIOSTTAKEFTVTIKHLQLNDAGQYVCQSGSDPTAEQN 237
QY 233 ADLQVLKPELVYEDLRGSVTFHCALGPEVANAFKFLCRQSGGNCVYVNTLGRAPA 292
Db 238 VDLRLL--TPGLLYGNLGGSVTFECALDSEDANAVASL-RQVRGNN--VVIDSQGTIDPA 292
QY 293 FEGRIILLNPODKGSFVITGLRKEDAGRYLCGAHSDGLOEGSPIQAWOLFVNEESTI 352
Db 293 FEGRIILFT-KAENGHFVSWTAGRKEDTGNLYLCGVNSGQSGDG-PTQLRQLFVNEIDV 350
QY 353 PRSPTVYVKGAGSVAVLCPYNRKESKIKYWCLEWGAQNGRCPLLDVDS-EGWYKAYEG 411
Db 351 SRSPVYLKGFPGGSVTIRCPYNPKRSDSHLQYLWESQTRH--LLVDSGEGLVKQDYTG 408
QY 412 RLSLLEEPGNGTFTVILNQLTSRDAGFYWCLTNGDILWRTVTEIKIIEGPNLKVGNVT 471
Db 409 RLALFEPPGNGTFTSVLNLQTLAEDEGFYWCVSDDESLLTSVKLQIVDGPSPFI-DKFT 467
QY 472 AVLGETLKVCHPCKFESSYKWKCKNNNTGCOALPSODEGSPKAFKNCNDSRLVSLTL 531
Db 468 AVQGEVPEITCHPCKFYSEKWKCKNNNDHGCEDLPTK-LSSSGDLVKCN-NNLVLTLT 525
QY 532 NLVTRADEGWYWCYKOGHYFGETAAYVAVEERKAAGSRDVSIAKADAAPD 583

```

Query Match 44.5%; Score 1435.5; DB 1; Length 773;
 Best Local Similarity 49.7%; Pred. No. 2.2e-102;
 Matches 316; Conservative 87; Mismatches 186; Indels 47; Gaps 18;

```

Db 526 DSVSEDEGWTCGAKDGHEFEVAAVRVELTEPAKVAVEPAKV-----PVDPAKAAAPA 581
QY 584 EKVLDS-----GFREIENK-----AIQDPLFAEZ 608
Db 582 EEKAKARCPVRRQWYPLSRKLTSCPEPLAAE 617

RESULT 7
ID R14670
AC R14670 standard; Protein; 584 AA.
DE 30-JAN-1992 (first entry)
DE Truncated poly Ig-receptor encoded by allele no. 1.
KW Rabbit; insemination; pregnancy.
OS Oryctolagus cuniculus.
FH Key Location/Qualifiers
FT peptide 1..18
FT domain /label= signal sequence
FT domain 10..118
FT domain /number= I
FT domain /note= "poly-Ig binding"
FT domain 119..223
FT domain /number= II
FT domain 224..332
FT domain /number= III
FT domain 333..441
FT domain /number= IV
FT domain 442..552
FT domain /number= V
FT domain 553..584
FT domain /number= IV
FT domain /note= "incomplete"
PN WO9116061-A.
PD 31-OCT-1991.
PF 16-APR-1991; U02604.
PR 16-APR-1990; US-510161.
PA (HARD ) HARVARD COLLEGE.
PI Kraehenbuhl JP, Weltzin RA, Neutra MR;
DR WPI: 91-339549/46.
DR N-PSDB; Q14498.
DE Stabilised poly-Ig complex contg. portion of poly-Ig receptor -
DE useful in protection against pathogens or against pregnancy
PS Disclosure; Fig 7; 51 pp; English.
CC The sequence was deduced from a cDNA clone of allele no. 1 and
CC is a truncated poly-Ig receptor. The native gene (Mostov et al)
CC is mutated to delete the portion encoding the transmembrane and
CC intra-cellular domains. The recombinant protein produced by
CC expression of the sequence is used as a stabiliser protein with a
CC poly-Ig specific for a selected antigen or family of antigens. The
CC compsn. can be administered directly to the mucosal surfaces of a
CC mammal to protect against a pathogen or against insemination. It
CC protects against allergens that contact the respiratory or digestive
CC mucosal surfaces and protects against pregnancy by cross-linking
CC sperm in the vagina.
CC See also R14671.
SQ Sequence 584 AA;

```

```

Query Match 44.4%; Score 1431; DB 1; Length 584;
Best Local Similarity 51.5%; Pred. No. 3.3e-102;
Matches 310; Conservative 81; Mismatches 175; Indels 36; Gaps 16;
QY 1 MLFLVLTCLLAVPAISTK-----SPIFGPEVNSVEGNSVITCYPTSYNRHTRKY 54
Db 1 MALFLVLTCLLAVSAATQAQSLGLGPSIFGPEVNVLEGSVITCYPTSYNRHTRK 60
QY 55 WCORGARGGCTITSSGYSVSSKYGAGRANLTNFPENGVVYVNIQAQLSDSGRYKCGLGI 114
Db 61 WCREEESGRCVTL-ASGYTSOEVSGRGKLTDFDKGEFVVYVDLTQNDSGSVKCGVG 119
QY 115 NSRGLSFDSVLEVSQSGPLLNDRVTYVDLGRVTINCPEKTKENAKRKSLSYKOIGLYPV 174

```

```

Db 120 NGRGLDFGVNVLVSQKPE--PDDVYVKQYESVTATTCPTTATROLKKSKFYKVEDGELY 177
QY 175 LVTDSSG--YVNPNTYGRJRLDIQGTGQLLFVYVNIQRLSDAGQYLCQAGDDSNKKN 232
Db 178 LIIDSSKKAQPRYKGRITLIQSTTAKEFTVIKHLQNDAGQIVCQSGSDPTAEQN 237
QY 233 ADQLVKPEPELYVEDLRGVSFTFHCALGPEVANAKFLCRQSGGNCNDVYVNTLGRAPA 292
Db 238 VDLRL--TPGLLYGNLGGSVTFECALDSEDANAVASL--RQVRGGN--VVIDSQGTIDPA 292
QY 293 FEGRIILNPQDKGSPSVVITGLRKEDAGRYLCCGHAHSDGLOEGSPQIAQWLFVNEESTI 352
Db 293 FEGRIILFT--KAENGHFVSVIAGLRKEDTNGYLCGVOSNGQSGDG--PTQLRQLFVNEEIDV 350
QY 353 PRSPYVKGAGSVVAVLCPYNNRKESIKYWCLEWGAQNGRCPLLVDS--EGWVKAQYEG 411
Db 351 SRSPVYKGFPGSGSVTIRCPYNPKRSDSHLQLYLWEGSOTRH--LLVDSGEGVLQKDYTG 408
QY 412 RLSLLEPENGFTVILNOLTSRDAGFYWCLNGDTLWRTTVEIKIIEGPNLKVPGNVT 471
Db 409 RLALFEEPNGTFSVYVNLQTLAEDEGFYWCVSDDDESLLTSVKLQIVDGEPSPTI--DKET 467
QY 472 AVLGETLKVPCHPCKFSSEYKYWCNNNTGCOALPQSDQEGSPKAFVNCDENSRLVSLTL 531
Db 468 AVQGEPEVITCHFPCKYFSEKYWCNNHGDGCDLPTK--LSSSGDLVKCN--NNLVLTLL 525
QY 532 NLVTRADEGWYCGVKQGHFYGETAAV-----YVAVEERKAAGSRDVSYLAKADAAPD 583
Db 526 DSVSEDEGWYCGAKDGHEFEVAAVRVELTEPAKVAVEPAKV-----PVDPAKAAAPA 581
QY 584 EK 585
Db 582 EE 583

RESULT 8
R14671
ID R14671 standard; Protein; 584 AA.
AC R14671;
DE 30-JAN-1992 (first entry)
DE Truncated poly Ig-receptor encoded by allele no. 2.
KW Rabbit; insemination; pregnancy.
OS Oryctolagus cuniculus.
FH Key Location/Qualifiers
FT peptide 1..18
FT domain /label= signal sequence
FT domain 10..118
FT domain /number= I
FT domain /note= "poly-Ig binding"
FT domain 119..223
FT domain /number= II
FT domain 224..332
FT domain /number= III
FT domain 333..441
FT domain /number= IV
FT domain 442..552
FT domain /number= V
FT domain 553..584
FT domain /number= IV
FT domain /note= "incomplete"
PN WO9116061-A.
PD 31-OCT-1991.
PF 16-APR-1991; U02604.
PR 16-APR-1990; US-510161.
PA (HARD ) HARVARD COLLEGE.
PI Kraehenbuhl JP, Weltzin RA, Neutra MR;
DR WPI: 91-339549/46.
DR N-PSDB; Q14499.
DE Stabilised poly-Ig complex contg. portion of poly-Ig receptor -
DE useful in protection against pathogens or against pregnancy
PS Disclosure; Fig 7; 51 pp; English.

```

CC The sequence was deduced from a cDNA clone of allele no. 2 and
CC is a truncated poly-Ig receptor. The native gene (Mostov et al)
CC is mutated to delete the portion encoding the transmembrane and
CC intra-cellular domains. The recombinant protein produced by
CC expression of the sequence is used as a stabilizer protein with a
CC poly-Ig specific for a selected antigen or family of antigens. The
CC compen. can be administered directly to the mucosal surfaces of a
CC mammal to protect against a pathogen or against insemination. It
CC protects against allergens that contact the respiratory or digestive
CC mucosal surfaces and protects against pregnancy by cross-linking
CC sperm in the vagina.
CC See also R14670.
CC Sequence 584 AA;
SQ

Query Match 38.7%; Score 1248; DB 1; Length 584;
Best Local Similarity 45.8%; Pred. No. 4e-88;
Matches 275; Conservative 78; Mismatches 215; Indels 32; Gaps 14;
QY 1 MLFVLCCLAVPAISTK-----SPIFGPEVNSVEGNSVITCYPTSVNRHTRKY 54
DB 1 MALFLLTCLLAVSAQAQSLGSPSIFGPEVNVLEGDSVITCYPTSVNRHSRKE 60
QY 55 WCQARGGCTILISSGYSKYAGRANLTNPENGTFFVYVNTAQLSDSGRYKCOLGI 114
DB 61 WCREESGRCVTL-ASTGYTTSQEYSGRNLTDFFPEKGEFVYVYLNRLSQSDSGYKCGVG 119
QY 115 NSRGLSDVLEVSQGPLLNDTRVYVLDGRTVINCPEKTENAKRKLXQIGLYPV 174
DB 120 NRGGLDFGVNVLVSQKXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX 179
QY 175 LVIDSSGYVNPNTGRIRLDIQGTQLLFVVINQLRSDAGQYLCOAGDSNSNKKND 234
DB 180 XXQLSDAGHYVCQGRNPTAEDQND 239
QY 235 LQVLKPELVYEDLRGVSFTHCALGPEVANVAKFLCRQSGENCVDVYVNTLGRAPAF 294
DB 240 LRL--TPDVLVNLGSGVTFECALDSEAKQVASL-RQVRGN--VVIDSQGTIDPAF 294
QY 295 GRILLNPQDQSGFSVITGLRKDAGRYLCGAHSDQLOEGSPQIAWLQVFNDEEIPR 354
DB 295 GRILFT-KAENGHSVVIAGLRDTONLYCGVNSQSGDG-PTQLWQVFNDEEIDYSR 352
QY 355 SPTVYKGVAGSSVAVLCPYNRKESKSIKWLWEGANGRCPLLVDS-EGWVQAQYEGRL 413
DB 353 SRPVLKGFPGSVTIRCPYNPKKSDSHQLYLWEGSQTRH--LLVDSGEGVLQKDYTGRL 410
QY 414 SLLEPENGFTVYVNLNLTSDAGFYWCLTNGDTLWRTTVEIKIEGEPNLKYPGNVAV 473
DB 411 ALFPEPENGFTSVYVNLNLTAEDEGFYWCVDSDDESLLTSYKLVQVDEGSPPTI-DEFTAV 469
QY 474 LGFTLVKVPCHFPKFFSYEKYCKWNNTGCOALPSODEGPKAFVNCDSRLVSLTLNL 533
DB 470 QGPEVETACHFPCKYFSEKYNCKWNHDGCEDLPTK-LSSGDLVKCN-NNLVLTLLDS 527
QY 534 VTRADEGWYKCGVQKHGYGETAAV-----YVAVEERKAAGSRDVSIAKADAAPDEK 585
DB 528 VSKDEGWYKCGAKDGHGFEFAVAVRVELTEPAKVAVEPAKV-----VVDPAKAAAPAAEE 583

RESULT 9

W99073
ID W99073 standard; Protein: 60 AA.
AC W99073;
DE Immunoglobulin like protein PolyIgRV4.
KW Human; PiGR-1; immunoglobulin; rheumatoid arthritis; psoriasis;
KW multiple sclerosis; systemic lupus erythematosus; diagnosis; receptor;
KW inflammatory bowel disease.
OS Unidentified.
PN EP-897981-A1.
PD 24-FEB-1999.
PF 11-AUG-1998; 306403.

PR 22-OCT-1997; US-955937.
PR 19-AUG-1997; US-056152.
PA (SMIK) SMITHLINE BEECHAM.
PI Hurle MR, Sweet RW, Truneh A, Wu S;
DR WPI: 99-134644/12.
PT New receptor (PiGR-1) polypeptide and polynucleotide - useful as
PT diagnostic reagents and for prevention and treatment of multiple
PT sclerosis, inflammatory bowel disease and psoriasis
PS Example 2; Page 14; 28pp; English.
CC The present invention describes a new receptor polypeptide designated
CC PiGR-1, which is a member of the immunoglobulin (Ig) superfamily. PiGR-1
CC proteins and polynucleotides are useful for diagnosing susceptibility to
CC diseases by detecting mutations in the PiGR-1 gene, and can diagnose
CC diseases associated with PiGR-1 protein imbalance by determining PiGR-1
CC protein expression levels. PiGR-1 proteins can be used to screen for
CC agonists and antagonists by measuring the binding to protein, and
CC observing the protein function. These can be used in treatment to
CC activate (agonist) or inhibit (antagonist) PiGR-1 activity, in addition
CC to direct administration of antisense sequences to prevent expression, or
CC PiGR-1 polynucleotides to treat conditions associated with a lack of
CC PiGR-1 protein. Gene therapy may also be used to affect endogenous PiGR-1
CC protein expression. PiGR-1 antibodies are useful for inducing an immune
CC response to immunise and prevent disease, and for isolating PiGR-1
CC clones or purifying the polypeptides by affinity chromatography. PiGR-1
CC proteins can be administered directly or as a vaccine to inoculate
CC against disease. Diseases diagnosed, prevented and treated include:
CC rheumatoid arthritis, multiple sclerosis, psoriasis, systemic lupus
CC erythematosus and inflammatory bowel disease. The PiGR-1 protein is also
CC useful for mapping the gene to a chromosome, allowing gene inheritance
CC to be studied through linkage analysis. The present sequence represents
CC an immunoglobulin like protein from the present invention.
SQ Sequence 60 AA;

Query Match 10.3%; Score 331; DB 1; Length 60;

Best Local Similarity 100.0%; Pred. No. 5.8e-19;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 353 PRSPTVYKGVAGSSVAVLCPYNRKESKSIKWLWEGANGRCPLLVDSGEGVQAQYEGR 412

DB 1 PRSPTVYKGVAGSSVAVLCPYNRKESKSIKWLWEGANGRCPLLVDSGEGVQAQYEGR 60

RESULT 10

W99072
ID W99072 standard; Protein: 57 AA.
AC W99072;
DE Immunoglobulin like protein PolyIgRV1.
KW Human; PiGR-1; immunoglobulin; rheumatoid arthritis; psoriasis;
KW multiple sclerosis; systemic lupus erythematosus; diagnosis; receptor;
KW inflammatory bowel disease.
OS Unidentified.
PN EP-897981-A1.
PD 24-FEB-1999.
PF 11-AUG-1998; 306403.
PR 22-OCT-1997; US-955937.
PR 19-AUG-1997; US-056152.
PA (SMIK) SMITHLINE BEECHAM.
PI Hurle MR, Sweet RW, Truneh A, Wu S;
DR WPI: 99-134644/12.
PT New receptor (PiGR-1) polypeptide and polynucleotide - useful as
PT diagnostic reagents and for prevention and treatment of multiple
PT sclerosis, inflammatory bowel disease and psoriasis
PS Example 2; Page 14; 28pp; English.
CC The present invention describes a new receptor polypeptide designated
CC PiGR-1, which is a member of the immunoglobulin (Ig) superfamily. PiGR-1
CC proteins and polynucleotides are useful for diagnosing susceptibility to
CC diseases by detecting mutations in the PiGR-1 gene, and can diagnose
CC diseases associated with PiGR-1 protein imbalance by determining PiGR-1
CC protein expression levels. PiGR-1 proteins can be used to screen for
CC agonists and antagonists by measuring the binding to protein, and
CC observing the protein function. These can be used in treatment to

CC activate (agonist) or inhibit (antagonist) PIGR-1 activity, in addition
CC to direct administration of antisense sequences to prevent expression, or
CC PIGR-1 polynucleotides to treat conditions associated with a lack of
CC PIGR-1 protein. Gene therapy may also be used to affect endogenous PIGR-1
CC protein expression. PIGR-1 antibodies are useful for inducing an immune
CC response to immunise and prevent disease, and for isolating PIGR-1
CC clones or purifying the polypeptides by affinity chromatography. PIGR-1
CC proteins can be administered directly or as a vaccine to inoculate
CC against disease. Diseases diagnosed, prevented and treated include:
CC rheumatoid arthritis, multiple sclerosis, psoriasis, systemic lupus
CC erythematosus and inflammatory bowel disease. The PIGR-1 protein is also
CC useful for mapping the gene to a chromosome, allowing gene inheritance
CC to be studied through linkage analysis. The present sequence represents
CC an immunoglobulin like protein from the present invention.
SQ Sequence 57 AA;

Query Match 8.9%; Score 287.5; DB 1; Length 57;
Best Local Similarity 95.0%; Pred. No. 1.2e-15;
Matches 57; Conservative 0; Mismatches 0; Indels 3; Gaps 2;

QY 22 IFGPEEYNSVEGNSVITCYPPPTSVNRHTRKYWCROGARGGCITLISSEGYVSSKYAGR 81

Db 1 IFGPEEYNSVEGNSVITCYPPPTSVN--TRKYWCROGAR--GCITLISSEGYVSSKYAGR 57

RESULT 11

W50033 ID W50033 standard; Protein; 532 AA.

AC W50033; DT 26-JUN-1998 (first entry)
DE Human immunity related factor.
KW Lymph node; human; immunity related factor; research; treatment;
KW immune disease; infectious disease.

OS Homo sapiens.

FH Key Location/Qualifiers

FT Peptide 1..16

FT Peptide /label= sig_peptide

FT Peptide 17..532

FT Peptide /label= mat_peptide

PN J10072495-A.

PD 17-MAR-1998.

PF 11-JUN-1997; 153218.

PR 13-JUN-1996; JP-152362.

PA (ASAH) ASAH KASEI KOGYO KK.

DR WPI: 98-234766/21.

DR N-PSDB: V20383

PT Immunity related factor - useful in the treatment of immune related

PT and infectious diseases

PS Claims 1, 3 and 4; Pages 18-20; 21pp; Japanese.

CC The present sequence is a lymph node derived human immunity

CC related factor, which can be used to research and treat immune and

CC infectious diseases.

SQ Sequence 532 AA;

Query Match 8.1%; Score 261.5; DB 1; Length 532;

Best Local Similarity 33.9%; Pred. No. 2.8e-12;

Matches 59; Conservative 22; Mismatches 50; Indels 43; Gaps 3;

QY 1 MLFLVFLCL-----LAVFPAISTKSPFGPEE----- 28

Db 1 MPULILCLGGSSFALPQKRPHRWLWEGSLPSTRHLRANGTLRPSPLCWREESFAA 60

QY 28 -----VNSVEGNSVITCYPPPTSVNRHTRKYWCROG--ARGGCITLISSEGYVSSK 77

Db 61 PNLKUGRLVSGEGGAVTIOCHYAPSVNRHTRKYWCRLGPPRWICQITVSTNQYTHR 120

QY 78 YAGRANTNPENFTFVYNTAQLSODSGRYKCGLGNRGLSFDVLSVQGP 131

Db 121 YRDRVALTDFPQGLFVRLSOLSPDDITGCVLCIGSENMLFLSMNLITISAGP 174

RESULT 12

W99078

ID W99078 standard; Protein; 43 AA.

AC W99078;

DT 13-MAY-1999 (first entry)

DE Immunoglobulin like protein PolyIgRV4.

KW Human; PIGR-1; immunoglobulin; rheumatoid arthritis; psoriasis;

KW multiple sclerosis; systemic lupus erythematosus; diagnosis; receptor;

KW inflammatory bowel disease.

OS Unidentified.

PN EP-897981-A1.

PD 24-FEB-1999.

PF 11-AUG-1998; 306403.

PR 22-OCT-1997; US-955937.

PR 19-AUG-1997; US-056152.

PA (SMIK) SMITHKLINE BEECHAM.

PI Hurle MR, Sweet RW, Truneh A, Wu S;

DR WPI: 99-134644/12.

DT New receptor (PIGR-1) polypeptide and polynucleotide - useful as

DT diagnostic reagents and for prevention and treatment of multiple

PT sclerosis, inflammatory bowel disease and psoriasis

PS Example 2; Page 14; 28pp; English.

CC The present invention describes a new receptor polypeptide designated

CC PIGR-1, which is a member of the immunoglobulin (Ig) superfamily. PIGR-1

CC proteins and polynucleotides are useful for diagnosing susceptibility to

CC diseases by detecting mutations in the PIGR-1 gene, and can diagnose

CC diseases associated with PIGR-1 protein imbalance by determining PIGR-1

CC protein expression levels. PIGR-1 proteins can be used to screen for

CC agonists and antagonists by measuring the binding to protein, and

CC observing the protein function. These can be used in treatment to

CC activate (agonist) or inhibit (antagonist) PIGR-1 activity, in addition

CC to direct administration of antisense sequences to prevent expression, or

CC PIGR-1 polynucleotides to treat conditions associated with a lack of

CC PIGR-1 protein. Gene therapy may also be used to affect endogenous PIGR-1

CC protein expression. PIGR-1 antibodies are useful for inducing an immune

CC response to immunise and prevent disease, and for isolating PIGR-1

CC clones or purifying the polypeptides by affinity chromatography. PIGR-1

CC proteins can be administered directly or as a vaccine to inoculate

CC against disease. Diseases diagnosed, prevented and treated include:

CC rheumatoid arthritis, multiple sclerosis, psoriasis, systemic lupus

CC erythematosus and inflammatory bowel disease. The PIGR-1 protein is also

CC useful for mapping the gene to a chromosome, allowing gene inheritance

CC to be studied through linkage analysis. The present sequence represents

CC an immunoglobulin like protein from the present invention.

SQ Sequence 43 AA;

Query Match 7.2%; Score 232; DB 1; Length 43;

Best Local Similarity 100.0%; Pred. No. 1.5e-11;

Matches 43; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 413 LSLLEPFGNGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEI 455

Db 1 LSLLEPFGNGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEI 43

RESULT 13

W99077

ID W99077 standard; Protein; 46 AA.

AC W99077;

DT 13-MAY-1999 (first entry)

DE Immunoglobulin like protein PolyIgRV1.

KW Human; PIGR-1; immunoglobulin; rheumatoid arthritis; psoriasis;

KW multiple sclerosis; systemic lupus erythematosus; diagnosis; receptor;

KW inflammatory bowel disease.

OS Unidentified.

PN EP-897981-A1.

PD 24-FEB-1999.

PF 11-AUG-1998; 306403.

PR 22-OCT-1997; US-955937.

PR 19-AUG-1997; US-056152.

PA (SMIK) SMITHKLINE BEECHAM.

PI Hurle MR, Sweet RW, Truneh A, Wu S;

DR WPI; 99-134644/12.
 PT New receptor (PIGR-1) polypeptide and polynucleotide - useful as
 PT diagnostic reagents and for prevention and treatment of multiple
 PT sclerosis, inflammatory bowel disease and psoriasis
 PS Example 2; Page 14; 28pp; English.
 CC The present invention describes a new receptor polypeptide designated
 CC PIGR-1, which is a member of the immunoglobulin (Ig) superfamily. PIGR-1
 CC proteins and polynucleotides are useful for diagnosing susceptibility to
 CC diseases by detecting mutations in the PIGR-1 gene, and can diagnose
 CC diseases associated with PIGR-1 protein imbalance by determining PIGR-1
 CC protein expression levels. PIGR-1 proteins can be used to screen for
 CC agonists and antagonists by measuring the binding to protein, and
 CC observing the protein function. These can be used in treatment to
 CC activate (agonist) or inhibit (antagonist) PIGR-1 activity, in addition
 CC to direct administration of antisense sequences to prevent expression, or
 CC PIGR-1 polynucleotides to treat conditions associated with a lack of
 CC PIGR-1 protein. Gene therapy may also be used to affect endogenous PIGR-1
 CC protein expression. PIGR-1 antibodies are useful for inducing an immune
 CC response to immunise and prevent disease, and for isolating PIGR-1
 CC clones or purifying the polypeptides by affinity chromatography. PIGR-1
 CC proteins can be administered directly or as a vaccine to inoculate
 CC against disease. Diseases diagnosed, prevented and treated include:
 CC rheumatoid arthritis, multiple sclerosis, psoriasis, systemic lupus
 CC erythematosus and inflammatory bowel disease. The PIGR-1 protein is also
 CC useful for mapping the gene to a chromosome, allowing gene inheritance
 CC to be studied through linkage analysis. The present sequence represents
 CC an immunoglobulin like protein from the present invention.
 SQ Sequence 46 AA;

Query Match 6.7%; Score 215; DB 1; Length 46;
 Best Local Similarity 91.3%; Pred. NO. 3.4e-10;
 Matches 42; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 82 ANLTNFPNGTFFVNIQLSDSDSRYKCGLGINSRGLSFDVSLEY 127
 DB 1 ANLTNFPNGTFFVNIQLSDSDSRYKCGLGINSRGLSFDVSLEY 46

RESULT 14
 W94269
 ID W94269 standard; protein; 592 AA.
 AC W94269;
 DT 26-APR-1999 (first entry)
 DE H6FXscfv(CEA6)tripBscfv(CEA6) fusion protein sequence.
 KW Trimeric polypeptide; tetranectin trimerising structural element; TTSE;
 KW fusion protein; ligand binding structure; toxin; enzyme; cytokine;
 KW artificial antibody; pharmacokinetic; pharmacodynamic; gene therapy;
 KW transfection; imaging; tumour; human; tetranectin; CEA6.
 OS Synthetic.
 OS Homo sapiens.
 PN WO9856906-A1.
 PD 17-DEC-1998.
 PF 11-JUN-1998; DK0245.
 PR 11-JUN-1997; DK-000685.
 PA (ETZE//) ETZERODT M.
 PA (GRAV//) GRAVERSEN N J H.
 PA (HOLT//) HOLTET T L.
 PA (KAST//) KASTRUP J S.
 FI Ezerodt M, Graversen NJH, Holtet TL, Kastrup JS,
 PI Larsen IK, Nielsen BB, Thøgersen HC;
 DR WPI; 99-080897/07.
 PT New monomer polypeptide constructs for diagnosis and therapy -
 PT comprise a tetranectin trimerising structural element covalently
 PT linked to at least one heterologous moiety for providing functional
 PT activity
 PS Example 4; Fig 20; 110pp; English.
 CC The invention relates to the design of trimeric polypeptides using
 CC polypeptide structural elements derived from the tetranectin protein
 CC family. The trimeric polypeptides constructed as a monomer polypeptide
 CC construct comprise at least one tetranectin trimerising structural
 CC element (TTSE) which is covalently linked to at least one heterologous
 CC moiety, the TTSE being capable of forming a stable complex with 2 other

CC TTSEs, with the proviso that the heterologous moiety is different from
 CC any of the fusion proteins CIIH6FXTN123, H6FXTN123, H6FXTN12, H6FXTN23
 CC (W94261 to W94264). The TTSE can be used for the construction of
 CC conjugates with heterologous moieties such as a ligand binding structure,
 CC a toxin, a detectable label, an in situ activatable substance, an enzyme,
 CC a radioactive moiety, a cytokine, a non-proteinaceous polymer, a photo
 CC cross-linking agent, or a group facilitating conjugation of the monomer
 CC polypeptide construct to a target. They can be used as vehicles for
 CC assembling antibody fragments into oligomeric or multivalent entities
 CC for generating chimeric artificial antibodies having preselected
 CC pharmacokinetic and/or pharmacodynamic properties. The constructs can
 CC be used for targeted gene therapy involving selective delivery of the
 CC material for transfection or infection of the specific population of
 CC cells. They can also be used for delivering a substance to a cell or
 CC tissue or for delivering an imaging or toxin-conjugated antibody to a
 CC tumour. They can also be used for prevention or treating a disease or
 CC for diagnosis. The TTSE provides a stable structure which can act as a
 CC vehicle for a wide variety of conjugates. The present sequence represents
 CC a H6FXscfv(CEA6)tripBscfv(CEA6) fusion protein sequence encoded by the
 CC plasmid pH6FXscfv(CEA6)tripBscfv(CEA6).
 SQ Sequence 592 AA;

Query Match 6.4%; Score 207.5; DB 1; Length 592;
 Best Local Similarity 19.2%; Pred. No. 4.7e-08;
 Matches 115; Conservative 74; Mismatches 196; Indels 215; Gaps 23;
 QY 33 GNSVSYTC---YYPPTSNNRHTKRYKCRQA-----RGGCITLISSEGVSSKVKAGR 81
 DB 31 GSSVKVCKASGGTFSNPSIN-----WLQAPQGLWVGSIIPSGTANY-AQKQGR 83
 QY 82 ANLTNFPNGTFFVNIQLSDSDSRYKCG-----LGINRGLSFDVS----- 125
 DB 84 LTITADESTAYMELSSLSASEDTAVYACGRSHNYELYYIMDVWGQGTMTVTSSSGGG 143
 QY 125 -----LEVSQPGLLNDTKYVYDGLRTVINCPEKTENAKRSLKYQIGLY 172
 DB 144 SGGGGSGGSDIQMTQSPSTL-----SASIGDRVITC-----RASEGIYHWLAWY 190
 QY 173 -----PVLVIDSSGVVNPYTGRIQLDIQGTQLLFSVVINQLRLSDAQYLCQ----- 222
 DB 191 QOKPGKAPKULLIYKASSLASGAPSRFSSGSGTD---FTLTISSLPDFFATYICQOYSN 247
 QY 222 -----AGDSSNNKKNADLQVLKPEPEL-----VYEDLR 250
 DB 248 YPLTFGGGTLEIKRAAAEQKLISEEDLNGAGTEPPTQPKKIYNAKKNADVVNTKMFELK 307
 QY 251 GSVTFHCALGPEVA-----NVAKFLCRQSSGENCDVYV 283
 DB 308 SRLD---TLADEVALLKEQALQTSQVQLQSGAEVKKPGSSVKVCKASGGTFSNPSI 364
 QY 284 NTLGKRAPA-----FEGRILLNPQDKGFSYVITGLRKEDAGR 322
 DB 365 NWL-RQAPGGGLEWVGSIIPSGTANYAQFGRLTITADESTAYMELSSLSASEDTAV 423
 QY 323 YLC-----GAHSDGQLQEGSPIQAWOLFVNEESTI 352
 DB 424 YACGRSHNYELYYIMDVWGQGTMTVTSSSGGGGSGGGGGSDIQMTQ----- 475
 QY 353 PRSTVYKGVAGSSVAVLCPYNRKESIKYWCLEWGAQNGRCPLLYDSEGWVAKQVEGR 412
 DB 475 --SPSTLSASIGDRVITC-----RASEGIYHWLAWYQOKPKAPKLIYK--ASLSAGA 526
 QY 413 LSLLEEPNGGT-FVYILNOLTSRDAGFYWC-----LTNGDTLWRTVTEIKIIEGPNL 464
 DB 527 PSRFGSGSGGDTFTLTISSLPDFFATYICQOYSNYPITFGG----GKLEIKRAAAEQKL 583

RESULT 15
 W42086
 ID W42086 standard; Protein; 1910 AA.
 AC W42086;
 DT 28-SEP-1998 (first entry)

Query Match	6.1%;	Score	195.5;	DB 1;	Length	1910;			
Best Local Similarity	19.3%;	Pred.	No. 2.1e-06;						
Matches	127;	Conservative	86;	Mismatches	240;	Indels	205;	Gaps	

QY	5	VLTCLLAVFPAITKSPIGPPEVINSVEG-----NSVSITCYYPPTSVNR--HRRKYWC	56						
	:	: : :	:	:	:	:	:	:	:
Db	43	LVPCPAAGIPVTLRWYLATGEIYDVGIRHVHPNGTQLQIFPPFPSSFTLIHDNTYYC	102						
	:	: : :	:	:	:	:	:	:	:
QY	57	-----ROGARGG-----CITLISSEGIVS-----	76						
	:	: : :	:	:	:	:	:	:	:
Db	103	TAEPSGKIRSODVHIKAVLRFPYTVRBDQTRMGNVAFCIIIPSSVEAYITVVSWEK	162						
	:	: : :	:	:	:	:	:	:	:
QY	76	---SKYAGRANLTNPENCTFFVNTAQLSQDDSGRYKC-----GLGINSRGLSFDVSL	125						
	:	: : :	:	:	:	:	:	:	:
Db	163	DTVSLVSGSRFL--ITSTGALYIKDVQ-NEDGLYNRCITRHRYTGETRQSARLFVSD	219						
	:	: : :	:	:	:	:	:	:	:
QY	126	EYSQGPELLNDIKVYTVDLGRVTINCPCPKTENAOKRKSLYKQIGLYPVLVIDSSGYNVP	185						
	:	: : :	:	:	:	:	:	:	:
Db	220	PANSASIILDGFDRHKAMAGORVEUPC-----KAL-----GHPEP	254						
	:	: : :	:	:	:	:	:	:	:
QY	186	NX-----TGRILDIOGTGOLLFSVINOLRLSDAGOYLCOAQDDDSNSKKNAAD	234						
	:	: : :	:	:	:	:	:	:	:
Db	255	DYRWLKDNMPLESLGRFQTIVG-----LLIEINRPDSGSYCEVSNRYGTAKVIGR	307						
	:	: : :	:	:	:	:	:	:	:
QY	235	LOYLKE-----EPELYIEDLRGSVTFHFCAIGPEVANVAKFLCRQSSGENCDVYVWNTLGRR	289						
	:	: : :	:	:	:	:	:	:	:
Db	308	LYVKQPCLKATIIPRKVKSSGVGSQVSLSCS-----VTGTDEQE	344						
	:	: : :	:	:	:	:	:	:	:
QY	290	APAFEGRIILLNFODKDGSFVSFITGLR-----KEDAGRYLCGAHSDG-----	332						
	:	: : :	:	:	:	:	:	:	:

Db 345 LSWYRNGEILNP-----GKNYRITGINHENLIMDHMYKSDGGAYQCFYRKDKLSAQDYVQ 399
Qy 332 -OLOEGSP--IOAWOLFVNEESTIPRSPTVVKGVAGSSVAVLCPYNRKESKIKYWCLWE 388
Db 400 VLEDGTPKIISAF-----SEKVVSPAEP-----VSLMCNVKGTPLPTIT-WTL-- 443
Qy 389 GAQNGRCPLLVDSEGWVKAQYEGRLSLEEPGNGTFTVILNQLTSRDAGFYWCLTN---G 445
Db 443 -----DDDPILKGGSHRISQMITSEGNVVSYLNISSQVRDGGVYRCTANNASAG 491
Qy 446 DTLARTTVEIKIIEGEPNLKYPGNVAVLGETLKVPCHPCKFSSEYKWKCKNNNTGCCOA 505
Db 492 VVLYQARINVR---GPASIRPMKNITAGR-----DTIHCVRVIGIPYISIKWYKNS-NL 543
Qy 506 LPSODEGPKAFVNCDENSRLVSLTLNLVTRADEGWYKGVKGHFGYGETAAVYVAVE 563
Db 544 LPFNHR--QVAF----ENNGLKLS-DYQKEVDEGEYTCNVLVQPOLSTOSVHVTVK 594

Search completed: November 20, 1999, 20:28:29
Job time: 28 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 1999, 21:06:24 : Search time 12.78 Seconds
(without alignments)
553.217 Million cell updates/sec

Title: US-09-095-385-4
Perfect score: 3226
Sequence: 1 MLLFVLTCLLAVFAIPAISTKS.....SGFREIENKAIQDPRLFAEZ 608

Scoring table: BLOSUM62

Searched: 122436 seqs, 11628473 residues

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
3: /cgn2_6/ptodata/1/iaa/PTUS9_COMB.pep.*
4: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1435.5	44.5	624	2	US-08-642-406A-22
2	184	5.7	483	2	US-08-392-338A-19
3	173.5	5.4	553	2	US-08-263-911-7
4	172.5	5.3	553	2	US-08-661-052-16
5	169.5	5.3	647	3	PCT-US93-00031-11
6	168.5	5.2	739	3	PCT-US93-00031-9
7	168.5	5.2	735	3	PCT-US93-00031-13
8	168.5	5.2	736	3	PCT-US93-00031-15
9	164	5.1	1021	1	US-08-497-025-3
10	163.5	5.1	828	1	US-08-261-304-7
11	163.5	5.1	662	1	US-08-261-304-7
12	161.5	5.0	596	2	US-08-752-307B-13
13	157.5	4.9	615	2	US-08-752-307B-9
14	153.5	4.8	647	3	PCT-US93-00031-23
15	151	4.7	1018	2	US-08-452-052-2
16	147.5	4.6	605	2	US-08-752-307B-8
17	147	4.6	642	1	US-08-217-299-1
18	147	4.6	734	2	US-08-389-459A-17
19	147	4.6	698	2	US-08-602-725-36
20	145	4.5	644	3	PCT-US93-00031-21
21	143.5	4.4	553	2	US-08-263-911-9
22	139.5	4.3	270	2	US-08-652-507-2
23	138.5	4.3	227	2	US-08-681-432-2
24	138.5	4.3	271	2	US-08-894-922A-10
25	137.5	4.3	243	1	US-08-230-843-4
26	137.5	4.3	243	2	US-08-636-936-4
27	137	4.2	1018	1	US-08-408-093-6
28	137	4.2	1018	1	US-08-408-420A-6
29	137	4.2	1018	1	US-08-714-901-6
30	137	4.2	879	1	US-08-554-612C-1
31	136.5	4.2	240	1	US-08-488-113B-148
32	136.5	4.2	240	2	US-08-477-848B-148
33	136.5	4.2	240	2	US-08-646-360-148
34	136.5	4.2	1447	3	PCT-US94-05277-2
35	136	4.2	643	3	PCT-US93-00031-19
36	135	4.2	252	2	US-08-894-922A-14
37	135	4.2	464	2	US-08-602-725-32
38	134.5	4.2	465	2	US-08-752-307B-5
39	134.5	4.2	462	2	US-08-752-307B-7

Sequence 17, Appl
Sequence 5, Appl
Sequence 16, Appl
Sequence 10, Appl
Sequence 14, Appl
Sequence 14, Appl

40 131.5 4.1 740 3 PCT-US93-00031-17
41 130 4.0 1311 2 US-08-340-011-5
42 127 3.9 622 2 US-08-356-786-16
43 126 3.9 611 2 US-08-752-307B-10
44 124.5 3.9 780 1 US-08-232-538-14
45 124.5 3.9 780 2 US-08-786-164-14

ALIGNMENTS

RESULT 1
US-08-642-406A-22
; Sequence 22, Application US/08642406A
; Patent No. 5959177
; GENERAL INFORMATION:
; APPLICANT: Hein, Mich B.
; APPLICANT: Hiatt, Andrew C.
; APPLICANT: Ma, Julian K.C.
; TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED
; TITLE OF INVENTION: SECRETORY ANTIBODIES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
; STREET: 10666 NO. 5959177th Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/642.406A
; APPLICATION NUMBER: US/08/642.406A
; FILING DATE: 03-MAY-1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/591,823
; FILING DATE: 02-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/427,765
; FILING DATE: 27-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Logan, April C.
; REGISTRATION NUMBER: 33,950
; REFERENCE/DOCKET NUMBER: 184.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 624 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-642-406A-22

Query Match 44.5%; Score 1435.5; DB 2; Length 624;
Best Local Similarity 49.7%; Pred. No. 3e-124;
Matches 316; Conservative 87; Mismatches 186; Indels 47; Gaps 18;
Qy 1 MLLFVLTCLLAVFAIPAISTK-----SPIFGPEEVNSVEGNSVITCYPTTSVNRHTRKY 54
Db 1 MALLFVLTCLLAVFAIPAISTK-----SPIFGPEEVNSVEGNSVITCYPTTSVNRHTRKY 60
Qy 55 WCRGARGGGITLISSEGGYSSKYAGRANITNFPENGTFFVYNIAQLSQDDSGRYKCGLGI 114
Db 61 WCREESGRCVTL-ASGYTSQEYSGRGKLTDFDKGEFVVYDQLTQDNGSGYKCGVG 119
Qy 115 NSRGLSFDVSLVSQSGPLGNDTKVYTVDLGRTVTINCPFKTENAKRKSILYKIGLYPV 174

```
Db 120 NGRGLDFGVNLVSKPE--PDDVVYKQYESYVTTTCPTTYATRLQKRSFYKVEGELY 177
QY 175 LVLDSSG--YVNPNTGRIRLDIOGTQQLFSVVINQLRLSDAGQLCQAGDDSNKKN 232
Db 178 LIIDSSKEAKDPYKGRITLQIOSTTAKEFTYIKHLQLNDAGQYVQSSDPTAEQN 237
QY 233 ADLQVLKPELVEDLURGSVTHCALGPEVANVAKFLCRQSGGENDVNVNLTGRAPA 292
Db 238 VDLRL--TPGLLYGNLGGSVTFECALDSEDANAVASL-RQVRGNN--VVIDSQGTIDPA 292
QY 293 FEGRIILLNPDQKGSFVWITGLRKEADAGRYLCGAHSDGOLQEGSPIOAQWLFFVNEESTI 352
Db 293 FEGRILET-KAENGHFSVVIAGLKEKEDTGNLGVQSGNGSGDG-PTQLRLQFLVNEEIDV 350
QY 353 PRSPTVVYGVAGSVAVLCPYNRKESKSIKYWCLMEGAQNGRCPLLVDS-EGWVKAQYEG 411
Db 351 SRPPLVKGFGGSGVTIRCPYNPKRSDSHLQYLVWEGSQTRH--LLVDSGEGLVQKDYTG 408
QY 412 RLSLLEPGNGTTVILNQLTSRDAGFYWCLTNGDILMRTTVEIKIEGEPNLUKVPNGVT 471
Db 409 RLALFEEPGNGTFSVNLNQLTAEDEGFYWCVDDESLSLTSVKLQIVDGEPSPTI-DKET 467
QY 472 AVLGELTKVPCHEPCKFESSYKWKYKWNNTGCOALPSQDEGSPKAFVNCDSRLVSLTL 531
Db 468 AVOGEPVEITCHEPCKFYFSSEKIYKWNHGDCELDPTK-LSSGDLVKCN-NNLVLTTL 525
QY 532 NLVTRADEGWYKGVKQGHFYGETAAV-----VXAVEERKAAGSRDVSIAKADAAPD 583
Db 526 DVSVEDEGWYKCGAKDGHGFEVAVRVVELTEPAKVAVEPAKV-----PVDPKAAAPAPA 581
QY 584 EKVLDS-----GPREIENK---AIQDRLFAEZ 608
Db 582 EERAKARCPVRRQWYPLSRKLRTSCPEPRLAAE 617
```

RESULT 2

```
US-08-392-338A-19
; Sequence 19, Application US/08392338A
; Patent No. 5869620
; GENERAL INFORMATION:
; APPLICANT: Whitlow, Marc
; APPLICANT: Wood, James F.
; APPLICANT: Hardman, Karl
; APPLICANT: Bird, Robert
; APPLICANT: Filpula, David
; TITLE OF INVENTION: Multivalent Antigen-Binding Proteins
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Avenue, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/392,338A
; FILING DATE: 22-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/989,846
; FILING DATE: 20-NOV-1992
; APPLICATION DATA:
; APPLICATION NUMBER: US 07/796,936
; FILING DATE: 25-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldstein, Jorge A.
; REGISTRATION NUMBER: 29,021
```

```
; REFERENCE/DOCKET NUMBER: 0977.00300007
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 483 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-392-338A-19

Query Match 5.7%; Score 184; DB 2; Length 483;
Best Local Similarity 22.2%; Pred. No. 8.4e-09;
Matches 109; Conservative 76; Mismatches 203; Indels 104; Gaps 24;

QY 127 VSQGPGLLNDTKVYTVDLGRVTINCP-----FKTENAKRKSLYKQ-IGLYPVLVI--- 178
Db 4 MSQSPSSL-----PVSYGKVTLSCKSSQSLLYSGNKNYLAWYQKPGOSPKLLIYWA 57
QY 178 --DSSGYVNPNTGRIRLDIOGTQQLFSVVINQLRLSDAGQYLCQ----- 222
Db 58 SARESG-VDPDRFTG-----SGSG-TDFTLSISSVKTTEDLAVIYCOQYYSPLTFGAGTK 109
QY 222 ---AGDDSNSNKN---ADLQVLKPEPELVYEDLRGSVTFHC-ALGPEVANVAKFLCRQS 274
Db 110 LVLKGSTSGSKSSBGKQVQLQSQDAELVKPG--ASVKISKASGYTFTDHAHIVKQ- 167
QY 275 SGENDCVVNLTGRKRA-----PEGRILLNPQDKGFSFVITGLRKEDAGRYLC 325
Db 167 ---NPEQGLEWIGYFSPGNDDFKYNEREKFKATLTADKSSSTAYVQLNSLTSEDSAVYFC 223
QY 326 GAHSDGQLQEGSPIOAQWLFFVNEESTIPRSPTVVYGVAGSVAVLCPYNRKESKSIKY-- 384
Db 224 -TRSLNAYWG---QGTSTVSSDVMSQSPSSLPVSVGEKVTLSG-----KSSQSLLSG 275
QY 384 ---WCLMEGAQNGRCPLLVDSSEGWVKAQYEGRLSLLEPGNGT-FTVILNQLTSRDAGF 438
Db 276 NKNYLAWYQKPGOSPKLLIY--WASARESGVDPDRFTGSGSGTDFTLSSISSVKTEDLAV 333
QY 439 YWC-----LTNGD-----TLWRTTVEIKIEGEPNLUKVPNGVTAVL--GETLKVPCHE 484
Db 334 YVCOQYYSPLTFGAGTKLVKLGSTSGSKSSBGKQVQLQSQDAELVKPGASVAKISCKA 393
QY 485 P-CKFSSYKWKYCKWN-----NTGCOALPSQDEGSPKAFVNCDSRLVSLTLN 532
Db 394 SGYFTFDHAHIVKQNPQGLEWIGYFSPGNDDFKYNEREKFKATLTADKSSSTAYVQLN 453
QY 533 LVTRADEGWYK 544
Db 454 SLTSEDSAVYFC 465
```

RESULT 3

```
US-08-263-911-7
; Sequence 7, Application US/08263911
; Patent No. 5877291
; GENERAL INFORMATION:
; APPLICANT: Mezes, Peter S
; APPLICANT: Goullie, Brian B
; TITLE OF INVENTION: MULTIVALENT SINGLE CHAIN ANTIBODIES
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Duane C. Ulmer
; STREET: P. O. Box 1967
; CITY: Midland
; STATE: MI
; COUNTRY: US
; ZIP: 48641-1967
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

```

US-08-661-052-16
; Sequence 16, Application US/08661052
; Patent No. 5837243
; GENERAL INFORMATION:
; APPLICANT: Yashwant M. Deo
; APPLICANT: Joel Goldstein
; APPLICANT: Robert Graziano
; APPLICANT: Chezian Somasundaram
; TITLE OF INVENTION: THERAPEUTIC COMPOUNDS COMPRISED
; TITLE OF INVENTION: OF ANTI-FC RECEPTOR ANTIBODIES
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD
; STREET: 60 State Street, Suite 510
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIA TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/661.052

```

PRIORITY: 1
 APPLICATION NUMBER: US 08/484,172
 FILING DATE: 07-JUNE-1995
 ATTORNEY/AGENT INFORMATION:
 NAME: Arnold, Beth E.
 REGISTRATION NUMBER: 35,430
 REFERENCE/DOCKET NUMBER: MXI-043CP
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (617)227-7400
 TELEFAX: (617)227-5941
 INFORMATION FOR SEQ ID NO: 16:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 553 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-661-052-16

[illegible]

4

```
QY 347 N-----FESTIPRPTVVKGAVSSAVLPCPNRKESKSIKYNCLWEGAQ 391
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 395 SSGGGGGGGGGGSENVLTPSPAINASPGKVTITC-----SASSSVSY-MHWFOOK 449
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 392 NGRCPLLVDSEG-----VKAQYGRLSLLEPPGNGT-FTVLNOLTSRDAGFYW 440
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 450 PGTSFKL-----WTYSNLSAGVPARFSG-----SGSGTSYLSLTSRWEADAATY 497
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 441 C-----LTNGDTLWRTTVEIKIEG 460
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 498 CQORSSYPLTFGA---GTKLELRAAG 521
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 5
PCT-US93-00031-11
; Sequence 11, Application PC/TUS9300031
; GENERAL INFORMATION:
; APPLICANT: Osborn, Laurelee
; APPLICANT: Benjamin, Christopher D.
; TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE FOURTH
; TITLE OF INVENTION: IMMUNOGLOBULIN-LIKE DOMAIN OF VCAM1
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive
; CITY: Chicago
; STATE: IL
; COUNTRY: US
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/00031
; FILING DATE: 19930112
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McNicholas, Janet M.
; REGISTRATION NUMBER: 32,918
; REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 715-1000
; TELEFAX: (312) 715-1234
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-00031-11

Query Match 5.3%; Score 169.5; DB 3; Length 647;
Best Local Similarity 21.3%; Pred. No. 2.9e-07;
Matches 138; Conservative 91; Mismatches 225; Indels 195; Gaps 36;

QY 33 GNSVSTICYPTTSVNRHTRKYNCRQARGCIT-LISSEGYSSYAGRANLTNPENG 91
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 40 GDSVSLTC-----STTGCSPPFSNRTQIDSPNGK--VTNEGTT 78
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 92 TFVNTAQLSQDSDGRYKCGLGINSRGLSFDVLSVQSGPGLNDTKY---TVDLGRV 148
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 79 TLTMPNVPFSGNEHS---YLTCATCESKLEKIQVEIYFSP---KDPFIHLSGPLEAGKPI 133
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 149 TINC-----PF-----KTENAK-----RKSLYK---OIGLYPVL----- 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 134 TVKCSADVVPFDRLEIDLKGDHLMKSEFFLEDADRKSLKSTVTPVTPVTDIGKVL 193
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 176 -----VIDSSGYVNPNT-----GRIRLDIQGTG----- 200
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 194 VCRAKLHIDEMDSVPVTRQAVKELQYVISPKNVTISVNPSTKLQEGSVTMTCSSEGLPA 253
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

```
QY 200 -QLLFS-----VVINQLRLSDAGQYLCOAGDDSNKKNADL--QVLKP 240
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 254 PEIFWSKKLDNGLQSLGNATLTLIAMMEDSGIYVCEGVNLICKRKEVELIQAAPR 313
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 241 EPELYVEDLRG-----SVTFHCALGPVAVNAKFLCRQSSGNCDDVVVNTLKGKRAPAF 294
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 314 DPEI---EMSGGLVNGSVTVSCVKV-PSVYPLDRLEIELLKE-----TILENIEFLE 362
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 295 GRILLNPQDKGFSFWITGLRKEDAGRYL-CGA--HSDGOLQEGSPIQAWQ-LFVNES 350
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 363 DTDMSLENKSLWTFIPT---IEDTKALVCOAKLHDDMEFEFKQSQSTQTLVN--- 417
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 351 TTPRSPTVVKGAVSSAVLPCPNRKESKSIKYNCLWEGAQNGRCPLLVDSGWWKAQYE 410
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 417 VAPRDTTVL--VSPSSIL-----EGSSVNMTCLSQGF--APKIL---WSRQLPN 460
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 411 GRLSLEPPGNGTFTVLNQLTSRDAGFYWCL-TNGDTLWRTTVEIKIEGEPNKLKVPGN 469
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 461 GELQPLSE--NATLTLSTKM--EDSGVYLCGEGINQAGRSRKEVELIIQVTPKDIK---- 513
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 470 VTAVLGETLKV--PCHFPCFSSYEKYWC---KWNNTCCQALPS-----QDEG 512
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 513 LTAPPSVSKEGDTVIISCTCGNVPETWILKKAETGDTVLKSIDGAYTIRKAOLKDAG 572
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 513 PSKAFVNCDENSRLVSLTLNLVTRADEGWYCGYKQGHFYGETAAVYVA 561
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 573 VYECESKNKVGSRSLTLDVQGRN-----NKDYFSPPELLVLYFA 613
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 6
PCT-US93-00031-9
; Sequence 9, Application PC/TUS9300031
; GENERAL INFORMATION:
; APPLICANT: Osborn, Laurelee
; APPLICANT: Benjamin, Christopher D.
; TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE FOURTH
; TITLE OF INVENTION: IMMUNOGLOBULIN-LIKE DOMAIN OF VCAM1
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive
; CITY: Chicago
; STATE: IL
; COUNTRY: US
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/00031
; FILING DATE: 19930112
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McNicholas, Janet M.
; REGISTRATION NUMBER: 32,918
; REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 715-1000
; TELEFAX: (312) 715-1234
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 739 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-00031-9

Query Match 5.2%; Score 168.5; DB 3; Length 739;
Best Local Similarity 24.1%; Pred. No. 4.5e-07;
```


Db 366 SLGDEBEGAYRCVVAEVMKTRTGS-----WQVLQRKQS---PDSHVHLRKPAAARSVVV----- 416
Qy 374 NRKESKIKYWCLEWEG-----AQNGRCPLLYD-----SEGHWK- 407
Db 416 ---STKNKQOVWEGETLAFCLCRAGGAESPLSVSWHHIPRDOTQPEFVAGMGODGIVQL 471
Qy 407 -----AQYEGRLSLEBPGNGTTVILNQLTSRDAGFYWCLTN-----GDTLNRITV 453
Db 472 GASYGVPSPYHGN--TRLEKMDWATQLEITFTAITDSGTIECRVSEKSRNOARDLSWTQKI 530
Qy 454 EIKIIEGPNLKV 466
Db 531 SVTVKSLESSQV 543

RESULT 10
US-08-261-304-2
; Sequence 2, Application US/08261304
; Patent No. 5708147
; GENERAL INFORMATION:
; APPLICANT: Cybulsky, Myron I.
; APPLICANT: Gimbrone, Michael A.
; APPLICANT: Collins, Tucker
; TITLE OF INVENTION: Mononuclear Leukocyte Directed
; TITLE OF INVENTION: Endothelial Adhesion Molecule Associated with
; TITLE OF INVENTION: Atherosclerosis
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Avenue, N.W.
; STREET: Suite 300
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: United States of America
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Ascii
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,304
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649,565
; FILING DATE: 01-FEB-1991
; APPLICATION NUMBER: U.S. 07/487,038
; FILING DATE: 02-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 0627.2100004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 833-7533
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 828 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-261-304-2

Query Match 5.1%; Score 163.5; DB 1; Length 828;
Best Local Similarity 24.5%; Pred. No. 1.6e-06;
Matches 137; Conservative 65; Mismatches 226; Indels 131; Gaps 31;
Qy 25 PEEVNSVEGNSVITCYPTPTSVNRHTRKYWCROGARGGCTTL-ISSEG-----YVSSKY 78
Db 208 PKERTTKELOVYISPNTVISVNPSTRL-----QSGGSVTWTCSEGLPVPEIFNSKK 261
Qy 79 AGRANLTFNPENGTFVNVNIAQLSODDSGRYKCGLGINSRGLS-FDVSLEVSQGFGLLNDT 137

Db 262 QDNGNQLRSGNAT---LTLIAMRMEDSGIYVCE-GYNQIGKSKREVELIVQERP----- 313
Qy 138 KVTVDLGRVTITNCFPKTEENAKRKSLYKQIGLYPVLVIDSSGYVNPNTVIGRILDIQ 197
Db 313 --FTVETSPGPRIAA-----QIGDPVLTCSVRGCETPSFWSRTQIDSPL 355
Qy 198 TGOLESFVWINQLRSDAG-----QYLCQAGDDSNKKNADLOVLK-----PEPELVYED 248
Db 356 NGQVTSEGTKSLTLSPVSFENEHSYLCVTGCHKKLEKGIQVELYSFPRDPEIELSGPP 415
Qy 249 LRG-SVTFHCALGPEVANVAKFLCROSSGENCDVVNTLGKRAPAFEGRILLNPQKDG- 307
Db 416 VNGRPVTVC-----KVPNYPF-----DRLIELLKGETMMKKEFLEEDKSL 461
Qy 307 -SFSVVTIGL-RKEDAGRYL-CGA---HSDGQLQEGSPIOAWO-LFYNEESTIPRSPTVVK 360
Db 462 ETASLEMTFIPMEDTGKVLVCOAKLHIDEMEFEPQORSTQPLFVN---VAPRD---IAV 516
Qy 361 GVAGSSVAVLCPYNRKESKIRYWCLEWGAQNGRCPLLYDSEGWKAQYEGRLSLEEPG 420
Db 517 WVSPPSIV-----EGRSVNMTCCSYGLP---APKIL---WSRLKNGDLQPLSE-- 561
Qy 421 NGFTTILNQLTSRDAGFYWCLTNGDTL---WRTTVEIKI-----TEGEPNLKVPGN 459
Db 561 NTLALISTKL--EDSGIYVC--EGINLAGSKRVELVIOVAPKDIQLTAFPSKSVKEG 616
Qy 470 VTAVLGETLKVCPCHFPCKFSSEYKWC---KWNNTCCQALPS-----QDEGPS 514
Db 617 DTVIIS-----CTCGNVPETWILKKAETGDTVLKSIDGAYTIRKAOLEDAGVY 666
Qy 515 KAFVNCDENSRLVSLTNL 533
Db 667 ECESKNEVGSQLRSLTDV 685
RESULT 11
US-08-261-304-7
; Sequence 7, Application US/08261304
; Patent No. 5708147
; GENERAL INFORMATION:
; APPLICANT: Cybulsky, Myron I.
; APPLICANT: Gimbrone, Michael A.
; APPLICANT: Collins, Tucker
; TITLE OF INVENTION: Mononuclear Leukocyte Directed
; TITLE OF INVENTION: Endothelial Adhesion Molecule Associated with
; TITLE OF INVENTION: Atherosclerosis
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox
; STREET: 1225 Connecticut Avenue, N.W.
; STREET: Suite 300
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: United States of America
; ZIP: 20036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Ascii
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/261,304
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/649,565
; FILING DATE: 01-FEB-1991
; APPLICATION NUMBER: U.S. 07/487,038
; FILING DATE: 02-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Potter, Jane E. R.
; REGISTRATION NUMBER: 33,332
; REFERENCE/DOCKET NUMBER: 0627.2100004
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 833-7533
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 828 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-261-304-7

REFERENCE/DOCKET NUMBER: 0627.2100004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 833-7533
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 662 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-361-304-7

Query Match 5.19; Score 163.5; DB 1; Length 662;
Best Local Similarity 24.39; Pred. No. 1.1e-06;
Matches 129; Conservative 72; Mismatches 214; Indels 115; Gaps 30;
QY 46 SVNRHTRKYMCRQAGGCITL-SSSEGYVS-----SKYAGRANLTNFPENGIFVNVIAQ 99
DB 205 VSNPSTKL-----QEGSVITMCSSEGLPAPEIFWKKLNDGNLQSLGNAT--LTLIA 256
QY 100 LSODDSGRYKCGLGINSRGLS-FDVSLEVSQPGLLNDTKVYTDLGRVTVINCPFKTEN 158
DB 257 MRWEDSGIYVCE-GVNLIGNRKEVELIVQEKP-----FTVEISPGPRIA----- 302
QY 159 AQKRKSLYKOIGLYPVLVDSSGVVNPNTGRIRLDIOGTGQLLFSVINOLRLSDAG-- 217
DB 302 -----QIGDSVMLTCSVMGCSPSFWRTOIDSPLSGKVRSEGTNSTLTLSPVSYFE 352
QY 217 --QYLC--QAGDDSNKKNADLOVLKPEPELYVEDLRG-----SVTFHCALGPEVAN 265
DB 353 NEHSILVTCVCGHKKLEKGLOVELYTPRPEI---EMSGLVNGSSVTSCKY-PSVYP 408
QY 266 VAFELCRQSSGENDVYVNTLGRAPAFEGRIILNPQDKGFSFVITGLRKEDAGRYL- 325
DB 409 LDRLEIELLAGE-----ILENIELEDTDMKSLNKSLEMTFTPT---IEDTCKALV 458
QY 325 CGA--HSDGLOQSGSPQAWQ-LFVNEESTIPRSPVTVKGVAGSSVAVLPYNRKSKSI 381
DB 459 CQAKLHIDDEFEFPKQSTQTLVYN--VAPRDTTVL--VSPSSIL-----EEGSSV 506
QY 382 KYCLWEGAQNGRCPLVDSEGVWKAQYEGRLSLLEPPGNGTFTVILNOLTSRDAGFYWC 441
DB 507 NMTCLSGGFP---APKIL-----NSRQLPNGELOPLSE--NATLTLSITKM--EDSGYILC 555
QY 442 L-TNGDTLWRTTVEIKIEGEPNLKVPNGYNTAVLGETLV--PCHEPKCFSSYEKYWC-- 497
DB 556 EGINAQGRSRKEVELIIQVTPKDIK-----LTAPPSSEVKEGDTVIISCTCGNVPETWIL 611
QY 497 -KNNTCCOALPS-----QDEGFSKAFVNCDENSRLVSLTLNL 533
DB 612 KKAETGDTVLKSIDGAYTIRKAQLKDAGYECESKNKVGSQLRSUTLDV 661

RESULT 12
US-08-752-307B-13
Sequence 13, Application US/08/52307B
Patent No. 5952171
GENERAL INFORMATION:
APPLICANT: McCarthy, Sean A.
APPLICANT: Gearing, David P.
APPLICANT: Levinson, Douglas A.
TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES
TITLE OF INVENTION: ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson, P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: US
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/752,307B
FILING DATE: 19-NOV-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: MeikieJohn, Ph.D., Anita L.
REGISTRATION NUMBER: 35,283
REFERENCE/DOCKET NUMBER: 09404/020001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-542-5070
TELEFAX: 617-542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 596 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-752-307B-13
Query Match 5.09; Score 161.5; DB 2; Length 596;
Best Local Similarity 21.59; Pred. No. 1.4e-06;
Matches 95; Conservative 72; Mismatches 155; Indels 119; Gaps 22;
QY 25 PEEVANSVEGNSYITCY---YPPTSVNRHTRKYMCRQAGGCITLISSEGVSSKYAGR 81
DB 233 PETLPAAGSTVKLECFALGNVPVQIN-----WRR-----SDG---MPFPK 271
QY 82 ANLTFPENGTVFNIAQLSQDSDGRYKCGLGINSRGLSFDVLSVQSGPGLNDTKVYT 141
DB 272 IKLRKF--NG--VLEIPNFQEDTGSYEC-IAENSRGKN-----VARG-RLTYAKPYW 319
QY 142 V----DLGRTVINPFKTEAKRSLYKQIGLYPVLVIDSSGVVNPNTGRIRLDIOG 197
DB 320 VOLLKDVETAVEDSLYWEGRASGKPKPSYRWLNKGDALVLEE-----RIQIEN 367
QY 198 TCOLFSVVINOLRLSDAGQYLCQAGDDSNKKNADLOVLKPEPELYVEDLRGVSFTFC 257
DB 368 G-----ALTIANLVSDSGMFOCIAENKHLIYSSAELKVLASAPDFSRLPMKKMIQV-- 421
QY 258 ALGPEVANYAKFLCRQSSGENC-----DVVVTNLGRAPAFEGRI-LLNPDQDKGSFS 309
DB 421 ----QVGLVILDCRPSASPRALSFWKKGDTVVRE-----QARISLNL-----DGGUK 464
QY 310 VVITGLRKEDAGRYLCGAHSDGLOQEGSPQAWQLFVNEESTIPRSPVTVKGVAGSSVAV 369
DB 465 IN--NVTKADAGIYTCIAEN---QFGRANGTTLVVTETRIILAPSNDVAVGESIIL 518
QY 370 LCPYNRKESKSIKYWCLWEGAQNGRCPLVDSEGVWKAQYEGRLSLLEP-----GN 421
DB 519 PC-----QVQHDPLLDIMFAN---YFNGTLTDFKDKDGHFKEVKGGS 556
QY 422 GTFTVILNOLTSRDAGFYWCL 442
DB 557 SSGDLMIIRNIQLKHGKVKYVCM 577
RESULT 13
US-08-752-307B-9
Sequence 9, Application US/08/52307B
Patent No. 5952171
GENERAL INFORMATION:
APPLICANT: McCarthy, Sean A.
APPLICANT: Gearing, David P.
APPLICANT: Levinson, Douglas A.
TITLE OF INVENTION: METHOD FOR IDENTIFYING GENES

```
; TITLE OF INVENTION: ENCODING NOVEL SECRETED OR MEMBRANE-ASSOCIATED PROTEIN
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: US
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/752,307B
; FILING DATE: 19-09-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Melkielejohn, Ph.D., Anita L.
; REGISTRATION NUMBER: 35,283
; REFERENCE/DOCKET NUMBER: 09404/020001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-542-5070
; TELEFAX: 617-542-8906
; TELEX: 200154
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 615 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-752-307B-9

; Query Match 4.9%; Score 157.5; DB 2; Length 615;
; Best Local Similarity 20.4%; Pred. No. 3.5e-06;
; Matches 114; Conservative 74; Mismatches 221; Indels 149; Gaps 27;

QY 28 VNSVEGNSVITCY----YPTSVNRHTRKRCQARGGCTLISSEGVSSKYAGRAN 83
DB 146 LEAVEGPEMLKCAAPDGPFPSTN-----WMQESIDGSIKINN-----SR 188

QY 84 LTNFPENGTFVYVNIAGLSQDSDGRYKCGLGINSRGLSFDVSLVSGPGLLNDTKVYTD 143
DB 189 MTLDPGNNLWFSNTRDASSDFYACSAITSVFRS-----EYKIGNKVLVDVKMGYS 241

QY 144 LGRTVTINCPFKTENAKRKSLY-KOIGLY-----PVLVIDSSGYVNPNTYGRILRD 194
DB 242 ASQNK--HPVRQYVSRQSALRGKRMELFCIYGGTPLPQTVMSKDGQ-RIQWSDRITQG 298

QY 195 IOGTGOLLSVVINQLRLSDAGOLYLCQAGDSDNSNKKNDLQVLKPEPELVYEDLRGVT 254
DB 299 HYCK-----SLVIRQTFNFDAGTYTCDVSGVNGNAQSFSLIINVSVPYFTKE----- 347

QY 255 FHCALGPEVANYAK-----FLCRQSGGENDCVVNTLTK-----RAPAFEGRIILLNPQDK 305
DB 347 -----PEIATAEDBEVVFECRAAGVPEPKTSWIHNGKPIEGSTP-----NPRRTV 392

QY 306 GSFVVITGLRREDAGRYLCGA-HSDGQLQEGSPIQAWOLFVN---EESTIPRSPTVVKG 361
DB 393 TQNTPIINLVKGDGTNYCGNATNSLGYVK-----DYVLNVQAEPTTISEAPAAVST 445

QY 362 VAGSSVAVLCPYNRKESKSIKYWCLWEGAQNGRCPLLDVSEGVYKQAEGRSLSLLEPGN 421
DB 446 VDGRRVNTIKRVNGSPKPLVK-----WLRASN-----WLTG-----GRYNV---QAN 484

QY 422 GTFTVILNOLTSRDAQFYWCLNGDTLWRTTVEIKIIEGPNLKV-----PGNWTAV 473
DB 485 GDLEI--QDVTFSADGKYTCYAGN-----KFGBIQADGSLVVKKEHTITQEPQNYEA 534
```

```
QY 474 LGETLKVPCHPCKFPSSVE-----KYWCKWNNTGCOALPQODEGPKSAFVNCDSNRL 526
DB 535 AQOSATFRCN-----EAHDDTLEIDWKMKGQSIDFEAQP-----RFVKTNDNSLT 581

QY 527 VSLTLNLVTRADEGHWYC 544
DB 582 IAKTMEL-----DSGEYTC 595
```

RESULT 14

```
PCT-US93-00031-23
; Sequence 23, Application PC/TUS9300031
; GENERAL INFORMATION:
; APPLICANT: Osborn, Laurelee
; APPLICANT: Benjamin, Christopher D.
; TITLE OF INVENTION: ANTIBODIES RECOGNIZING THE FOURTH
; TITLE OF INVENTION: IMMUNOGLOBULIN-LIKE DOMAIN OF VCAM1
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Allegrretti & Witcoff, Ltd.
; STREET: 10 South Wacker Drive
; CITY: Chicago
; STATE: IL
; COUNTRY: US
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/00031
; FILING DATE: 19930112
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: McNicholas, Janet M.
; REGISTRATION NUMBER: 32,918
; REFERENCE/DOCKET NUMBER: 92,306-A; D001 CIP PCT
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 715-1000
; TELEFAX: (312) 715-1234
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US93-00031-23
```

```
Query Match 4.8%; Score 153.5; DB 3; Length 647;
Best Local Similarity 20.8%; Pred. No. 8.8e-06;
Matches 141; Conservative 101; Mismatches 237; Indels 199; Gaps 39;
```

```
QY 5 VLTCLAVFPAISTK-SPIFGPEEVSNEGVSVITCYPPPTSVNRHTRKRCQARGG 63
DB 14 ILWIMFAASQAFTVEISP--GPRIAAQI-GDSVMLTC-----SVN-----G 51

QY 64 CIT-LISSEGVSSYKAGRANLTNFPENGTFVYVNTAQISQDSDGRYKCGLGINSRGLSFD 122
DB 52 CBSPSFSVRTQIDSPLSGKVR---SEGTNSTLTLSPVSEFENEHSYLTCTVTCGHKKLEK 107

QY 123 VSLEVSQGPGLLNDPKY---TVDLGRVTITNC-----PF-----KTENACK----- 162
DB 108 LOVELYSFP---KDPEIHLSPLEAGKPIYKCSVADVYPFDRLEIDLKGDHLKMSOBF 164

QY 162 -----RKSLYK---QIGLYPVL-----VIDSSGYVNP 186
DB 165 LEDADRKSLKSLVETFTPTVIEDIGKVLVCRALKHIDEMDSVPTVROAVKELQVYISPK 224

QY 187 YT-----GRIRLDIQGTG-----QLLFS-----VVINOLRLS 213
```


GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 1999, 23:17:44 ; Search time 24.27 Seconds
(without alignments)
1600.573 Million cell updates/sec

Title: US-09-095-385-4
Perfect score: 3226
Sequence: 1 MLLFVLTCLLAVFPAISTKS.....SGFREIENKAIQDPLRAEZ 608

Scoring table: BLOSUM62

Searched: 504389 seqs, 63891290 residues

Database : Pending_Patents_AA:*

1: /cgn2_6/ptodata/1/paa/PCUR99_COMB.pep.*
2: /cgn2_6/ptodata/1/paa/US07_COMB.pep.*
3: /cgn2_6/ptodata/1/paa/US080_COMB.pep.*
4: /cgn2_6/ptodata/1/paa/US081_COMB.pep.*
5: /cgn2_6/ptodata/1/paa/US082_COMB.pep.*
6: /cgn2_6/ptodata/1/paa/US083_COMB.pep.*
7: /cgn2_6/ptodata/1/paa/US084A_COMB.pep.*
8: /cgn2_6/ptodata/1/paa/US084B_COMB.pep.*
9: /cgn2_6/ptodata/1/paa/US085_COMB.pep.*
10: /cgn2_6/ptodata/1/paa/US086_COMB.pep.*
11: /cgn2_6/ptodata/1/paa/US087_COMB.pep.*
12: /cgn2_6/ptodata/1/paa/US088_COMB.pep.*
13: /cgn2_6/ptodata/1/paa/US089_COMB.pep.*
14: /cgn2_6/ptodata/1/paa/US090_COMB.pep.*
15: /cgn2_6/ptodata/1/paa/US091_COMB.pep.*
16: /cgn2_6/ptodata/1/paa/US092_COMB.pep.*
17: /cgn2_6/ptodata/1/paa/US093_COMB.pep.*
18: /cgn2_6/ptodata/1/paa/US094_COMB.pep.*
19: /cgn2_6/ptodata/1/paa/US095_COMB.pep.*
20: /cgn2_6/ptodata/2/paa/PCT_NEW_COMB.pep.*
21: /cgn2_6/ptodata/2/paa/US07_NEW_COMB.pep.*
22: /cgn2_6/ptodata/2/paa/US08_NEW_COMB.pep.*
23: /cgn2_6/ptodata/2/paa/US09_NEW_COMB.pep.*
24: /cgn2_6/ptodata/2/paa/US60_NEW_COMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3226	100.0	764	12	US-08-860-037-2
2	3226	100.0	608	14	Sequence 2, Appli
3	3226	100.0	764	17	Sequence 4, Appli
4	3219	99.8	607	16	Sequence 30, Appli
5	3139	97.3	746	7	Sequence 8, Appli
6	2177	67.5	757	7	Sequence 4, Appli
7	2099.5	65.1	771	7	Sequence 6, Appli
8	2064.5	64.0	769	7	Sequence 8, Appli
9	1435.5	44.5	773	7	Sequence 10, Appli
10	1435.5	44.5	624	16	Sequence 2, Appli
11	557	17.3	106	13	Sequence 22, Appli
12	532.5	16.3	109	13	Sequence 6, Appli
13	528.5	16.4	102	13	Sequence 9, Appli
14	385	11.9	79	19	Sequence 7, Appli
15	331	10.3	60	13	Sequence 5172, Ap
16	331	10.3	60	17	Sequence 6, Appli
17	287.5	8.9	57	17	Sequence 6, Appli
18	287.5	8.9	57	17	Sequence 5, Appli
19	232	7.2	43	13	Sequence 5, Appli

20	232	7.2	43	17	US-08-300-985-11	Sequence 11, Appli
21	215	6.7	46	13	US-08-955-937A-10	Sequence 10, Appli
22	215	6.7	46	17	US-08-300-985-10	Sequence 10, Appli
23	195.5	6.1	1910	13	US-08-956-991-2	Sequence 2, Appli
24	195.5	6.1	1571	13	US-08-956-991-11	Sequence 11, Appli
25	187	5.8	224	17	US-08-316-412-29	Sequence 29, Appli
26	184	5.7	483	15	US-08-166-093-19	Sequence 19, Appli
27	184	5.7	483	15	US-08-166-094-19	Sequence 19, Appli
28	184	5.7	483	15	US-08-166-750-19	Sequence 19, Appli
29	184	5.7	483	15	US-08-172-019-19	Sequence 19, Appli
30	183	5.7	483	2	US-07-989-846B-19	Sequence 19, Appli
31	183	5.7	483	6	US-08-392-338-19	Sequence 19, Appli
32	182	5.6	390	13	US-08-961-564A-2	Sequence 2, Appli
33	177.5	5.5	211	19	US-60-127-852-8	Sequence 8, Appli
34	173.5	5.4	553	2	US-07-990-263-7	Sequence 7, Appli
35	172.5	5.3	553	17	US-09-364-088-16	Sequence 16, Appli
36	171	5.3	1461	9	US-08-564-963-2	Sequence 2, Appli
37	171	5.3	1461	9	US-08-564-963B-2	Sequence 2, Appli
38	170.5	5.3	613	19	US-60-127-852-1	Sequence 1, Appli
39	169.5	5.3	647	2	US-07-821-712-11	Sequence 11, Appli
40	169.5	5.3	647	6	US-08-342-642-5	Sequence 5, Appli
41	169.5	5.3	647	8	US-08-473-764-5	Sequence 5, Appli
42	169.5	5.3	647	8	US-08-477-463-5	Sequence 5, Appli
43	169.5	5.3	647	8	US-08-480-843-5	Sequence 5, Appli
44	169.5	5.3	647	8	US-08-484-383-5	Sequence 5, Appli
45	169.5	5.3	647	8	US-08-486-336-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-860-037-2
Sequence 2, Application US/08860037
GENERAL INFORMATION:
APPLICANT: Jan van Oostrum
APPLICANT: Fredericus A.M. Asselbergs
APPLICANT: Christian Leist
APPLICANT: Markus Grutter
TITLE OF INVENTION: Production of Recombinant Secretary
TITLE OF INVENTION: Component
NUMBER OF SEQUENCES: 7
CORRESPONDENCE ADDRESS:
ADDRESSEE: Michael Glynn, Novartis Corporation,
STREET: 59 Route 10
CITY: East Hanover
STATE: New Jersey
COUNTRY: U.S.
ZIP: 07936

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,037
FILING DATE: 11 June 1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/EP95/04797
FILING DATE: 12 June 1995
ATTORNEY/AGENT INFORMATION:
NAME: Furman, Diane E.
REGISTRATION NUMBER: 31,104
REFERENCE/DOCKET NUMBER: 4-20259/A/PCT
TELEPHONE: 908-277-2311
TELEFAX: 908-277-4009
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 764 amino acids
TYPE: amino acid
TOPOLOGY: linear

Db 601 DPRLFAEZ 608

RESULT 3

```

US-09-316-412-30
? Sequence 30, Application US/09316412
? GENERAL INFORMATION:
? APPLICANT: Sheppard, Paul O.
? TITLE OF INVENTION: IMMUNOMODULATOR POLYPEPTIDE 2SIG57
? FILE REFERENCE: 98-23
? CURRENT APPLICATION NUMBER: US/09/316,412
? CURRENT FILING DATE: 1999-05-20
? EARLIER APPLICATION NUMBER: 60/089,784
? EARLIER FILING DATE: 1998-06-18
? NUMBER OF SEQ ID NOS: 30
? SOFTWARE: FastSEQ for Windows Version 3.0
? SEQ ID NO 30
? LENGTH: 764
? TYPE: PRT
? ORGANISM: Homo sapiens
US-09-316-412-30

```

Query Match	100.0%;	Score 3226;	DB 17;	Length 764;
Best Local Similarity	99.8%;	pred. No. 2.6e-274;		
Matches 607;	Conservative 1;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	MLLFVLTCLLAVFAIPATSKSIFGPEEVNSVEGNSVITCYPPPTSVAHRHTRKYYWCROGA	60	
Db	1	MLLFVLTCLLAVFAIPATSKSIFGPEEVNSVEGNSVITCYPPPTSVAHRHTRKYYWCROGA	60	
Qy	61	RGGCITLISSEGYVSSKYAGRANLTNFPENGTFFVYNIAQLSODDSGRYKCGLGINSRGLS	120	
Db	61	RGGCITLISSEGYVSSKYAGRANLTNFPENGTFFVYNIAQLSODDSGRYKCGLGINSRGLS	120	
Qy	121	FDVLSLEYSQGGLLNDTKVYTVDLGRVTINCIPKFTENAKRKSLYKQIGLYPVLVIDSS	180	
Db	121	FDVLSLEYSQGGLLNDTKVYTVDLGRVTINCIPKFTENAKRKSLYKQIGLYPVLVIDSS	180	
Qy	181	GYVNPNTGTRIRLDIOGTGQLLFSVYINOLRLSDAGQYLCQAGDDSNSNKKNADLQVLKP	240	
Db	181	GYVNPNTGTRIRLDIOGTGQLLFSVYINOLRLSDAGQYLCQAGDDSNSNKKNADLQVLKP	240	
Qy	241	EPELVYEDLRGVSVTFHCALGPEVANVAKFLCRQSSGENCDVYVNTLGRAPAFEGRIILN	300	
Db	241	EPELVYEDLRGVSVTFHCALGPEVANVAKFLCRQSSGENCDVYVNTLGRAPAFEGRIILN	300	
Qy	301	PODKGFSFVITGLRKEDAGRYLCGAHSDQLOEGSPIQAWOLFVNEESTIPRSPTVVK	360	
Db	301	PODKGFSFVITGLRKEDAGRYLCGAHSDQLOEGSPIQAWOLFVNEESTIPRSPTVVK	360	
Qy	361	GVAGSSVAVLCPYNRKESKSIKYWCLMEGAQNGRCPLLVDSSEGWYKAYQEGRLSLLEPG	420	
Db	361	GVAGSSVAVLCPYNRKESKSIKYWCLMEGAQNGRCPLLVDSSEGWYKAYQEGRLSLLEPG	420	
Qy	421	NGTFTVILNQLTSDRAGFYWCLTNGDILWRTTVEIKIIEGEPNLKVPQNVYAVLGETLKY	480	
Db	421	NGTFTVILNQLTSDRAGFYWCLTNGDILWRTTVEIKIIEGEPNLKVPQNVYAVLGETLKY	480	
Qy	481	PCHEPCKAFSSYEKYWCKWNNTGCCALPSQDGEPSKAFVNCNDSRLVSLTLNLVTRADEG	540	
Db	481	PCHEPCKAFSSYEKYWCKWNNTGCCALPSQDGEPSKAFVNCNDSRLVSLTLNLVTRADEG	540	
Qy	541	WYWCVGKHGFYGETAAVYVAVEERKAAGSRDVSJAKADAAPEDKVLDSGFPREIENKAIQ	600	
Db	541	WYWCVGKHGFYGETAAVYVAVEERKAAGSRDVSJAKADAAPEDKVLDSGFPREIENKAIQ	600	

RESULT 4

```

US-09-275-667-8
; Sequence 8, Application US/09275667
; GENERAL INFORMATION:
; APPLICANT: Wycoff, Keith L.
; APPLICANT: Jaiswal, Sudhir K.
; TITLE OF INVENTION: SIMULTANEOUS INTRODUCTION OF MULTIPLE
; TITLE OF INVENTION: HETEROLOGOUS GENES INTO PLANTS AND ASSEMBLY INTO FUNCTIONAL
; TITLE OF INVENTION: MULTIMERS
; FILE REFERENCE: 415142000100
; CURRENT APPLICATION NUMBER: US/09/275,667
; CURRENT FILING DATE: 1999-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: fastseq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 607
; TYPE: prt
; ORGANISM: Homo sapiens
US-09-275-667-8

```

Query Match.	99.98;	Score 3219;	DB 16;	Length 607;	
Best Local Similarity	99.98;	Pred. No. 7.7e-274;			
Matches	606;	Conservative	1;	Mismatches 0; Indels 0; Gaps 0;	
Qy	1	MLFVLVCLLAVPPAISTKSP	IFCPPEEVNSVEGNSV	ITCYIPPTSVNRRHKYWCROGA 60	
Db	1	MLFVLVCLLAVPPAISTKSP	IFCPPEEVNSVEGNSV	ITCYIPPTSVNRRHKYWCROGA 60	
Qy	61	RGCGITILISSEGYVSSKYAGRANUTNP	PENGTFFVNTIAQLSQDSDS	GRYKCOLGINSRGLS 120	
Db	61	RGCGITILISSEGYVSSKYAGRANUTNP	PENGTFFVNTIAQLSQDSDS	GRYKCOLGINSRGLS 120	
Qy	121	FDVSLEYSQSGPLNDTKVTV	DLGRTVTINCPFKTENAKR	KSLYKOIGLYPVLVIDSS 180	
Db	121	FDVSLEYSQSGPLNDTKVTV	DLGRTVTINCPFKTENAKR	KSLYKOIGLYPVLVIDSS 180	
Qy	181	GYNPNTTGRIRLDIOGTG	QLLFSVWLNQLRSDAGOY	LCQAGDDSNKKNADQLVKP 240	
Db	181	GYNPNTTGRIRLDIOGTG	QLLFSVWLNQLRSDAGOY	LCQAGDDSNKKNADQLVKP 240	
Qy	241	EPELVYEDLKGSVTFH	CGALGPEVANVAKFLC	ROSSGNCDDVVNTLGKRAPAFEGRILLN 300	
Db	241	EPELVYEDLKGSVTFH	CGALGPEVANVAKFLC	ROSSGNCDDVVNTLGKRAPAFEGRILLN 300	
Qy	301	PODKGGSFVVITGLR	KEDAGRYLCGAHSDQ	LOEGSPIQAWQLFVNEESTIPRSPTVVK 360	
Db	301	PODKGGSFVVITGLR	KEDAGRYLCGAHSDQ	LOEGSPIQAWQLFVNEESTIPRSPTVVK 360	
Qy	361	GVAGSSVAVLCPYNRK	BESKSIKYWCLMEGAQ	NGRCPLLVDSEGWVKAQYEGRLSLLBEPG 420	
Db	361	GVAGSSVAVLCPYNRK	BESKSIKYWCLMEGAQ	NGRCPLLVDSEGWVKAQYEGRLSLLBEPG 420	
Qy	421	NGFTFTVLNQLTSR	DAGFYWCLNGDPLWRTT	VEIKIIEGEPNLKVPGNVTAVLGTELKV 480	
Db	421	NGFTFTVLNQLTSR	DAGFYWCLNGDPLWRTT	VEIKIIEGEPNLKVPGNVTAVLGTELKV 480	
Qy	481	PCHPCKFSSEYKVC	WKNNTGQALPSQDE	GPSSKAFVNCDENSR	LSVLTUNLVTRADEG 540
Db	481	PCHPCKFSSEYKVC	WKNNTGQALPSQDE	GPSSKAFVNCDENSR	LSVLTUNLVTRADEG 540
Qy	541	WYWCVKVQGHFYG	ETAAYVYVAEERKA	AGSRDYSLAKADAA	PDEKVLDSGFREIENKAIQ 600
Db	541	WYWCVKVQGHFYG	ETAAYVYVAEERKA	AGSRDYSLAKADAA	PDEKVLDSGFREIENKAIQ 600
Qy	601	DPLRF	AE 607		
Db	601	DPLRF	AE 607		

RESULT 5
US-08-434-000A-4
; Sequence 4, Application US/08434000A
; GENERAL INFORMATION:

APPLICANT: ANDREW C. HIATT, JULIAN
APPLICANT: K.-C. MA, THOMAS LEHNER
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Suite 4700
STATE: Los Angeles
COUNTRY: California
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
CURRENT APPLICATION DATA:
FILING DATE: US/08/434,000A
CLASSIFICATION: 435
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 1
APPLICATION NUMBER: 08/367,395
FILING DATE: 12/30/94
ATTORNEY/AGENT INFORMATION:
NAME: Guise, Jeffrey W.
REGISTRATION NUMBER: 34,613
REFERENCE/DOCKET NUMBER: 212/127
TELEPHONE: (619) 552-8400
TELEFAX: (619) 552-0159
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 746 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
TOPOLOGY: DESCRIPTION: Human Polyimmunoglobulin Receptor
US-08-434-000A-4

Query Match 97.3%; Score 3139; DB 7; Length 746;
Best Local Similarity 99.8%; Pred. No. 1.le-266;
Matches 589; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 19 KPIFGPEVNSVEGNSVITCYPTSVNRTKRYWCQGGAGCITLISSEGVSSKY 78
DB 1 KSPIFGPEVNSVEGNSVITCYPTSVNRTKRYWCQGGAGCITLISSEGVSSKY 60
QY 79 AGRANLTNPENGTFVNVIAQLSQDSDSGRYKGLGINSRGLSFDVSLVSQGPGLNDTK 138
DB 61 AGRANLTNPENGTFVNVIAQLSQDSDSGRYKGLGINSRGLSFDVSLVSQGPGLNDTK 120
QY 139 VYVDLGRVTINCPEKTEAKRSLYKQIGLYPVLVIDSSGYVNPNTGRIIDIOGT 198
DB 121 VYVDLGRVTINCPEKTEAKRSLYKQIGLYPVLVIDSSGYVNPNTGRIIDIOGT 180
QY 199 GOLLSFVNQLRLSDAGQYLCQAGDSSNKKNDLQVLPKPELVYEDLRGVSFTFCA 258
DB 181 GOLLSFVNQLRLSDAGQYLCQAGDSSNKKNDLQVLPKPELVYEDLRGVSFTFCA 240
QY 259 LGPEVANVAKFLCRSSGNGCDVYNTLGRAPAFEGRILLNPQDKDGSFVITGLRKE 318
DB 241 LGPEVANVAKFLCRSSGNGCDVYNTLGRAPAFEGRILLNPQDKDGSFVITGLRKE 300
QY 319 DAGRYLCGAHSDGQLOEGSPIQAWOLFVNEESTIPRPTVYKGVAGSSVAVLCPYNRKES 378
DB 319 DAGRYLCGAHSDGQLOEGSPIQAWOLFVNEESTIPRPTVYKGVAGSSVAVLCPYNRKES 360

DB 301 DAGRYLCGAHSDGQLOEGSPIQAWOLFVNEESTIPRPTVYKGVAGSSVAVLCPYNRKES 360
QY 379 KSIKYWCLWEAQRCPCLLVDSGKWAQYEGRLSLLEPGNGTFTVILNQLTSRDAGF 438
DB 361 KSIKYWCLWEAQRCPCLLVDSGKWAQYEGRLSLLEPGNGTFTVILNQLTSRDAGF 420
QY 439 YWCLTNGDTLWRTTVEIKIIEGEPNLKVPGNVAVLGETLKVPCFCHFPCKFSSYKWKW 498
DB 421 YWCLTNGDTLWRTTVEIKIIEGEPNLKVPGNVAVLGETLKVPCFCHFPCKFSSYKWKW 480
QY 499 NNTGQALPSODEGSPKAFVNCNDSRLVSLTLNLVTRADEGWYWCQKGFHFGETAAY 558
DB 481 NNTGQALPSODEGSPKAFVNCNDSRLVSLTLNLVTRADEGWYWCQKGFHFGETAAY 540
QY 559 YVAVEERKAAGSRDVSYLAADAPDEKVLDSGFREIENKAIQDPRLFAEZ 608
DB 541 YVAVEERKAAGSRDVSYLAADAPDEKVLDSGFREIENKAIQDPRLFAEZ 590

RESULT 6
US-08-434-000A-6
Sequence 6, Application US/08434000A
GENERAL INFORMATION:
APPLICANT: ANDREW C. HIATT, JULIAN
APPLICANT: K.-C. MA, THOMAS LEHNER
TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: Lyon & Lyon
STREET: 633 West Fifth Street
CITY: Suite 4700
STATE: Los Angeles
COUNTRY: California
ZIP: 90071
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
COMPUTER: IBM Compatible
OPERATING SYSTEM: IBM P.C. DOS 5.0
SOFTWARE: Word Perfect 5.1
CURRENT APPLICATION DATA:
FILING DATE: US/08/434,000A
CLASSIFICATION: 435
PRIOR APPLICATION DATA: including application
PRIOR APPLICATION DATA: described below: 1
APPLICATION NUMBER: 08/367,395
FILING DATE: 12/30/94
ATTORNEY/AGENT INFORMATION:
NAME: Guise, Jeffrey W.
REGISTRATION NUMBER: 34,613
REFERENCE/DOCKET NUMBER: 212/127
TELEPHONE: (619) 552-8400
TELEFAX: (619) 552-0159
TELEX: 67-3510
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 757 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
TOPOLOGY: DESCRIPTION: Bovine Polyimmunoglobulin Receptor
US-08-434-000A-6

Query Match 67.5%; Score 2177; DB 7; Length 757;
Best Local Similarity 67.2%; Pred. No. 3.le-182;
Matches 405; Conservative 81; Mismatches 113; Indels 4; Gaps 3;

QY 6 LTCLLAVPAITKSPIFGPEEVNSVEGNSVITCYPTSVNRHTRKYWCQARGGCI 65
Db 6 LACLLAIFPVVSMKSPIFGPEEVNSVEGRSVIKCYPTSVNRHTRKYWCQAGRCI 65
QY 66 TLISSEGVSSKYAGRANLTNPENGTFVNNIAQLSQDSDGKYKGLGINSRGLSFDVSL 125
Db 66 TLISSEGVSSDYGRANLTNPESGTFVVDISHUTHKDSGRKYKGLGINSRGLNFDVSL 125
QY 126 EYSQPGLLNDKTRVTDLGRVTTCNCFKTEAOKRSKSLYKQIGLYPLVLDSSGYVNP 185
Db 126 EYSQPAQASHAVTIDLGRVTTCNCFTRANSEKRSKSLCKTQDCFOVVDSTGYVSN 185
QY 186 NYTGRIIDIOGTGLLFSWVNLRLSDAGQYLCQAGDSDNSKKNADLQVLKPEPELV 245
Db 186 SKYRAHISILCTNLVFSVINRKLSDAGMYVCOAGDDAKADKINIDLQVLEPELV 245
QY 246 YEDLRGVTFFHCALGPEVANVAKFLCRSSGNCDCVNTLGRAPAFEGRILLNPQDK 305
Db 246 YGDLRSSVTDCSLGPEVANVPKFLCOKKNGGACNVNTLGRKKAQDFQGRIVSVPKD-N 304
QY 306 GSFVVTIGLRKEDAGRYLCGAHSDGQLEGSPICAWOLFVNEESTIPRSPTVYKGVAGS 365
Db 306 GSFVHTSLRKEDAGRYVCGAPEGEPDGPMPVQAWOLFVNEETAIPASPSVVKGVRG 364
QY 366 SVAVLCPYNRKESIKYKWLMEGAQRCPLLDVSEGVKAYQEGRLSLLEPEGNGTFT 425
Db 366 SVTVSCPYNPKDANSKAYKWEAQAQRCPLRVSRGLMKEQYEGRLVLTPEPNGTFT 424
QY 426 VILNOLTRDAGFYWCLNGDFTLWRTTVEIKIEGPNLKVPGNVTAVLGETLKVPCHP 485
Db 426 VILNOLTRDAGFYWCLNGDFTLWRTTVEIKIEGPNLKVPGNVTAVLGETLKVPCHP 484
QY 486 KFSSEYKWKNNKTCOALPSODEGSKAFVNCDSRLVSLTLNLTTRADEGWYWG 545
Db 486 KFSSEYKWKNNKTCOALPSODEGSKAFVNCDSRLVSLTLNLTTRADEGWYWG 544
QY 546 VKQGFYGETAAVYVAEERKAAGSRDYSKADAAAPDEKVLDSGFREIENKAIQDPLF 605
Db 546 VKQGFYGETAAVYVAEERKAAGSRDYSKADAAAPDEKVLDSGFREIENKAIQDPLF 601
QY 606 AEZ 608
Db 606 AKE 604

RESULT 7

US-08-434-000A-8
; Sequence 8, Application US/08434000A
; GENERAL INFORMATION:
; APPLICANT: ANDREW C. HIATT, JULIAN
; APPLICANT: K.-C. MA, THOMAS LEHNER
; TITLE OF INVENTION: IMMUNOGLOBULINS CONTAINING PROTECTION
; TITLE OF INVENTION: PROTEINS IN PLANTS AND THEIR USES
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: Word Perfect 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/434.000A
; FILING DATE:
; CLASSIFICATION: 435

; PRIOR APPLICATION DATA: including application
; PRIOR APPLICATION DATA: described below: 1
; APPLICATION NUMBER: 08/367,395
; FILING DATE: 12/30/94
; ATTORNEY/AGENT INFORMATION:
; NAME: Guise, Jeffrey W.
; REGISTRATION NUMBER: 34,613
; REFERENCE/DOCKET NUMBER: 212/127
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 552-8400
; TELEFAX: (619) 552-0159
; TELEX: 67-3510
; TELEX: SEQUENCE LISTING
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 771 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; TOPOLOGY: DESCRIPTION: Mouse Polyimmunoglobulin Receptor
; US-08-434-000A-8

Query Match 65.1%; Score 2099.5; DB 7; Length 771;
Best Local Similarity 64.2%; Pred. No. 2e-175;
Matches 395; Conservative 81; Mismatches 130; Indels 9; Gaps 5;

QY 1 MLLFVLVLCCLAVPAITKSPIFGPEEVNSVEGNSVITCYPTSVNRHTRKYWCROGA 60
Db 1 MRLYFTLLTVFSGVSTKSPIFGPEEVNSVEGNSVITCYPTSVNRHTRKYWCROGA 60
QY 61 RGGCITLISSEGVSSKYAGRANLTNPENGTFVNNIAQLSQDSDGKYKGLGINSRGLS 120
Db 61 SGMCTLISSEGVSSKYAGRANLTNPENGTFVNNIAQLSQDSDGKYKGLGINSRGLS 120
QY 121 FDSVLEYSQPGLLNDKTRVTDLGRVTTCNCFKTEAOKRSKSLYKQIGLYPLVLDSS 180
Db 121 FDSVLEYSQVPELPSDTHVTKIDGRNVTIECPKRENVPSKSLCKKTNQSCVLVIDST 180
QY 181 GYVNPNTGRILDIOGTGLLFSVNLRLSDAGQYLCQAGDSDNSKKNADLQVLKP 240
Db 181 EKNVPSYIGRAKLFMKGTDLTVFYVNIHSHLTHNDAGLYICOAGEGPSADKKNVDLQVLAP 240
QY 241 EPELVYEDLRGVTFFHCALGPEVANVAKFLCRSSGNCDCVNTLGRAPAFEGRILLN 300
Db 241 EPELLYKDLRSSVTFFCDLGRVANEAKYLCRMNK-ETCDVLIINTLGRKDDPDEGRILIT 299
QY 301 PQDKGGSFVVTIGLRKEDAGRYLCGAHSDGQLEGSPICAWOLFVNEESTIPRSPTVVK 360
Db 301 PKDDNGRFSVLITGLRKEDAGHYOCGAHSSGLPOEGWPIQTWOLFVNEESTIPRRSVVK 359
QY 361 GVAGSSVAVLCPYNRKESIKYKWLMEGAQRCPLLDVSEGVKAYQEGRLSLLPEPG 420
Db 361 GVTGGSVAIACPNPKESSSLKYKWRGEGDNGCHPCALVGTQAOVQOPEYEGRLALFDOPG 419
QY 421 NGFTVILNOLTRDAGFYWCLNGDFTLWRTTVEIKIEG--EPNLKY-PGNVTAVLGET 477
Db 420 NGTVVILNOLTTEDAGFYWCLNGDSRWRTTIELQVAEATREPNELEVTPOATAVLGET 479
QY 478 LKVPCHFPCKFSSEYKWKNNKTCOALPSODEGSKAFVNCDSRLVSLTLNLTTRA 537
Db 480 FTVSCHYPCPKFYSOEYKWKNNKCHILPSHDEGARQSSVSCDQSSQLVSMTLNPSKE 539
QY 538 DEGWYKVGKGFYGETAAVYVAEERKAAGSRDVS---SLAKADAAPEKVLDSGFRE 593
Db 540 DEGWYKVGKGFYGETAAVYVAEERKAAGSRDVS---SLAKADAAPEKVLDSGFRE 598
QY 594 IENKAIQDPRLFAEZ 608
Db 599 KENKAIQDPRLFAEZ 613


```
;
; TOPOLOGY: linear
; TOPOLOGY: DESCRIPTION: Rabbit polyimmunoglobulin receptor
US-08-434-000A-2

Query Match 44.5%; Score 1435.5; DB 7; Length 773;
Best Local Similarity 49.7%; Pred. No. 4e-117;
Matches 316; Conservative 87; Mismatches 186; Indels 47; Gaps 18;

Qy 1 MLFLVLTCLLAVFAISTK-----SPIFGPEEVNSVEGNSVITCYPPPTSVNRHTRKY 54
Db 1 MALFLTCLLAVFAISATAQSSLLGPSSIFGPEVNVLEGDVSITCYPTTSVTRHSRKF 60

Qy 55 WCRQARGGCITLISSEGIVSSKYAGRANLNFPENGTFVYVNIQAOLSDSGRYKCGLGI 114
Db 61 WCREESGRCVTL-ASTGYTSQEYSGRGLTDFPDGGEFVYVTDLTQNDSGSKCGVGV 119

Qy 115 NSRGLSFDVSLVSGPGLLNDTKVYVLDLGRVTINCPEKTENAKRKSLYKQIGLYPV 174
Db 120 NGRGLDFGVNVLVSQKPE--PDDVYKQYESYVTTTCPTTYATRLKKSFKYKVEDGELV 177

Qy 175 LVIDSSG--YVNPNTYGRIRLDIQTGQLLFSVYVNIQLRSDAGQYLCQAGDDSNKKN 232
Db 178 LIIDSSKEAKDPYKGRITLIQSTTAKEFTVTKHLQNDAGQYVCSGSDPTAEQN 237

Qy 233 ADLOVLKPEPELVYEDLRGVSFTFHCALGPEVANVAKFLCRSGGENDCVVNTLGRAPA 292
Db 238 VDLRLL--TPGLLYGNLGGSVTFECALDSEDANAVASL-RQVRGN--VVIDSQGTIDPA 292

Qy 293 FEGRIILNPQDKGSFVITGLRKEDAGRYLCGAHSDGOLQEGSPIOAWOLFVNEESTI 352
Db 293 FEGRIILFT-KAENGHFSVVIAGLRKEDTGNLYCGVNSGSGDG-PTQLRLQFVNEIDV 350

Qy 353 PRSPTVYKGVAGSSVAVLCPYNNRKSISIKYWCWLEGAQNGRCPLLYDS-EGWVKAQYEG 411
Db 351 SRSPVVLKGFPGGVSITRCPYNPRSDSHLQYLWEGSQTRH--LLVDSGGLYQKDYTG 408

Qy 412 RLSLLEPENGTFVILNQLTSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVGNVT 471
Db 409 RLALFEPEGNGTFSVYLNQLTAEDGEGFYWCVSDDDSLTTSVKLQIVDGEPSPI-DKFT 467

Qy 472 AVLGETLKVPCHEPCKFSSEYKWKWNNTGCOALPSQDEGPKAFVNCNDSRLVSLTL 531
Db 468 AVQGEPEVITCHFPCKYFSSKYWKWNHGDCEDLPTK-LSSSGDLVKCN--NNLVLTLL 525

Qy 532 NLVTRADEGWMYCGYKQGHFYGETAAV-----YVAVERKAAGSRDVSIAKADAAPD 583
Db 526 DSVSEDEGWMYCGAKDGHEFEVAAVRVELTEPAKVAVEPAKV----PVDPAKAAAPA 581

Qy 584 EKVLDS-----GFREIENK---AQDPRLFAE2 608
Db 582 EEKAKRCPVRRRWYPLSRKRLTSCPEPLLAEE 617

RESULT 10
US-09-200-657-22
; Sequence 22, Application US/09200657
; GENERAL INFORMATION:
; APPLICANT: Hehn, Mich B.
; APPLICANT: Hiatt, Andrew C.
; APPLICANT: Ma, Julian K.C.
; TITLE OF INVENTION: TRANSGENIC PLANTS EXPRESSING ASSEMBLED
; TITLE OF INVENTION: SECRETORY ANTIBODIES
; NUMBER OF SEQUENCES: 26
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THE SCRIPPS RESEARCH INSTITUTE
; STREET: 10666 North Torrey Pines Road, TPC-8
; CITY: La Jolla
; STATE: California
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```
;
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/200,657
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/642,406
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/427,765
; FILING DATE: 27-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Logan, April C.
; REGISTRATION NUMBER: 33,950
; REFERENCE/DOCKET NUMBER: 184.2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 554-2937
; TELEFAX: (619) 554-6312
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 624 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-200-657-22

Query Match 44.5%; Score 1435.5; DB 16; Length 624;
Best Local Similarity 49.7%; Pred. No. 2.9e-117;
Matches 316; Conservative 87; Mismatches 186; Indels 47; Gaps 18;

Qy 1 MLFLVLTCLLAVFAISTK-----SPIFGPEEVNSVEGNSVITCYPPPTSVNRHTRKY 54
Db 1 MALFLTCLLAVFAISATAQSSLLGPSSIFGPEVNVLEGDVSITCYPTTSVTRHSRKF 60

Qy 55 WCRQARGGCITLISSEGIVSSKYAGRANLNFPENGTFVYVNIQAOLSDSGRYKCGLGI 114
Db 61 WCREESGRCVTL-ASTGYTSQEYSGRGLTDFPDGGEFVYVTDLTQNDSGSKCGVGV 119

Qy 115 NSRGLSFDVSLVSGPGLLNDTKVYVLDLGRVTINCPEKTENAKRKSLYKQIGLYPV 174
Db 120 NGRGLDFGVNVLVSQKPE--PDDVYKQYESYVTTTCPTTYATRLKKSFKYKVEDGELV 177

Qy 175 LVIDSSG--YVNPNTYGRIRLDIQTGQLLFSVYVNIQLRSDAGQYLCQAGDDSNKKN 232
Db 178 LIIDSSKEAKDPYKGRITLIQSTTAKEFTVTKHLQNDAGQYVCSGSDPTAEQN 237

Qy 233 ADLOVLKPEPELVYEDLRGVSFTFHCALGPEVANVAKFLCRSGGENDCVVNTLGRAPA 292
Db 238 VDLRLL--TPGLLYGNLGGSVTFECALDSEDANAVASL-RQVRGN--VVIDSQGTIDPA 292

Qy 293 FEGRIILNPQDKGSFVITGLRKEDAGRYLCGAHSDGOLQEGSPIOAWOLFVNEESTI 352
Db 293 FEGRIILFT-KAENGHFSVVIAGLRKEDTGNLYCGVNSGSGDG-PTQLRLQFVNEIDV 350

Qy 353 PRSPTVYKGVAGSSVAVLCPYNNRKSISIKYWCWLEGAQNGRCPLLYDS-EGWVKAQYEG 411
Db 351 SRSPVVLKGFPGGVSITRCPYNPRSDSHLQYLWEGSQTRH--LLVDSGGLYQKDYTG 408

Qy 412 RLSLLEPENGTFVILNQLTSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVGNVT 471
Db 409 RLALFEPEGNGTFSVYLNQLTAEDGEGFYWCVSDDDSLTTSVKLQIVDGEPSPI-DKFT 467

Qy 472 AVLGETLKVPCHEPCKFSSEYKWKWNNTGCOALPSQDEGPKAFVNCNDSRLVSLTL 531
Db 468 AVQGEPEVITCHFPCKYFSSKYWKWNHGDCEDLPTK-LSSSGDLVKCN--NNLVLTLL 525

Qy 532 NLVTRADEGWMYCGYKQGHFYGETAAV-----YVAVERKAAGSRDVSIAKADAAPD 583
Db 526 DSVSEDEGWMYCGAKDGHEFEVAAVRVELTEPAKVAVEPAKV----PVDPAKAAAPA 581
```

```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-407-0700
; TELEFAX: 610-407-0701
; TELEX: 846169
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 109 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-961-564A-9

Query Match 16.5%; Score 532.5; DB 13; Length 109;
Best Local Similarity 97.2%; Pred. No. 4.1e-39;
Matches 106; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

QY 22 IFGPEEVSNEGVSITCYPPPTSVNRHTRKYWCRO-GARGG-CITLISSEGYVSSKYA 79
Db 1 IFGPEEVSNEGVSITCYPPPTSVNRHTRKYWCROPGARGGGLCITLISSEGYVSSKYA 60

QY 80 GRANLTNFPENGTFVNVIAQLSQDSDSGRYKCGLGINS-RGLSFDVSLEV 127
Db 61 GRANLTNFPENGTFVNVIAQLSQDSDSGRYKCGLGINSRLGLSFDVSLEV 109

RESULT 13
US-08-976-293-7
; Sequence 7, Application US/08976293A
; GENERAL INFORMATION:
; APPLICANT: WU, SHUJIAN
; APPLICANT: SWEET, RAYMOND
; APPLICANT: TRUNEH, ALEMSEGED
; TITLE OF INVENTION: PIGR-2 A MEMBER OF IMMUNOGLOBULIN GENE SUPERFAMILY
; FILE REFERENCE: GH-70238
; CURRENT APPLICATION NUMBER: US/08/976,293A
; CURRENT FILING DATE: 1997-11-21
; EARLIER APPLICATION NUMBER: 60/056,774
; EARLIER FILING DATE: 1997-08-25
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Homo sapiens
US-08-976-293-7

Query Match 16.4%; Score 528.5; DB 13; Length 102;
Best Local Similarity 95.1%; Pred. No. 8.3e-39;
Matches 98; Conservative 3; Mismatches 1; Indels 1; Gaps 1;

QY 353 PRSPTVYKGVAGSSVAVLCPYNRKESIKYWCLEGAQNGRCPLLYDSGWWKQAQYEGR 412
Db 1 PRSPTVYKGVAGSSVAVLCPYNRKESIKYWCLEGAQNGRCPLLYDSGWWKQAQYEGR 59

QY 413 LSLLEEPNGTFTVILNQLTSRDAGFYWCCLTNGDTLWRTTVEI 455
Db 60 LSLLEDPNGTFTVIMNQLTSRDAGFYWCCLTNGDTLWRTTVEI 102

RESULT 14
US-60-160-203-5172
; Sequence 5172, Application US/60160203
; GENERAL INFORMATION:
; APPLICANT: BONAZZI, VIVIEN
; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS AND
; TITLE OF INVENTION: USES THEREOF
; FILE REFERENCE: CLO00116
; CURRENT APPLICATION NUMBER: US/60/160,203
; CURRENT FILING DATE: 1999-10-19

```

Query Match 10.3%; Score 331; DB 13; Length 60;
Best Local Similarity 100.0%; Pred. No. 8.3e-22;
Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 353 PRSPTVVKGVAGSSVAVLCPPYNRKESKSIKYWCLWEGAQNGRCPLLDVSEGWYKAYEGR 412
DB 1 PRSPTVVKGVAGSSVAVLCPPYNRKESKSIKYWCLWEGAQNGRCPLLDVSEGWYKAYEGR 60
Search completed: November 20, 1999, 23:20:39
Job time: 175 sec

NUMBER OF SEQ ID NOS: 6374
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5172
LENGTH: 79
TYPE: PRT
ORGANISM: HUMAN
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(79)
OTHER INFORMATION: Xaa = Any Amino Acid
US-60-160-203-5172

Query Match 11.9%; Score 385; DB 19; Length 79;
Best Local Similarity 92.4%; Pred. No. 2.3e-26;
Matches 73; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
QY 9 LLAVFPAISTKSPIFGPEEVNSVITCYPTSVNRHTRKYWCROGARGGCITLI 68
DB 1 LSSVPTAISTKSPIFGPEEVNSVIXCYPTSVNRHTRKYWCROGARGGCITLI 60
QY 69 SSEGYVSSKYAGRANLTNF 87
DB 61 SSEGYVSSKYAGRANLTNF 79

RESULT 15
US-08-955-937A-6
Sequence 6 Application US/08955937A
GENERAL INFORMATION:
APPLICANT: WU, SHUJIAN
APPLICANT: SWEET, RAYMOND
APPLICANT: TRUNEH, ALEMSEGED
TITLE OF INVENTION: PIGR-1, A MEMBER OF IMMUNOGLOBULIN
TITLE OF INVENTION: GENE SUPERFAMILY
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: RATNER & PRESTIA
STREET: P.O. BOX 980
CITY: VALLEY FORGE
STATE: PA
COUNTRY: USA
ZIP: 19482
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/955,937A
FILING DATE: 17-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/056,152
FILING DATE: 19-AUG-1997
ATTORNEY/AGENT INFORMATION:
NAME: PRESTIA, PAUL F
REGISTRATION NUMBER: 23,031
REFERENCE/DOCKET NUMBER: GH-70228
TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-407-0700
TELEFAX: 610-407-0701
TELEX: 846169
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 60 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-955-937A-6

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 1999, 21:54:53 ; Search time 14.27 Seconds
(without alignments)
1707.065 Million cell updates/sec

Title: US-09-095-385-4
Perfect score: 3226
Sequence: 1 MLLFVTLCLLAVFAISTKS.....SGFREIENKAIQDPRLEFAEZ 608
Scoring table: BLOSUM62
Searched: 122810 seqs, 40065486 residues
Database : PIR60:*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3226	100.0	764	1 QRHUGS	secretory componen
2	2178	67.5	757	2 I45956	polymERIC immunogl
3	2177	67.5	757	2 S48841	secretory componen
4	2064.5	64.0	769	1 QRRTGS	secretory componen
5	1435.5	44.5	773	1 QRRBGS	secretory componen
6	210.5	6.5	1323	2 PNO568	connectin 3B - chi
7	197	6.1	7962	2 I38346	elastic titin - hu
8	187	5.8	224	2 I37243	CdkRf-35 antigen -
9	185	5.7	1028	2 I38164	BIG-1 protein - ra
10	182.5	5.7	4391	2 A38096	perlecan precursor
11	173.5	5.4	664	2 I56171	B-cell adhesion pr
12	170	5.3	1427	2 I51669	tumor suppressor -
13	169.5	5.3	647	2 B41288	vascular cell adhe
14	169	5.2	1694	2 S50065	slalodhesin - mou
15	168.5	5.2	847	2 JH0371	B-cell adhesion pr
16	168.5	5.2	739	2 A41288	vascular cell adhe
17	167.5	5.2	1028	2 A53449	plasmacytoma-assoc
18	165	5.1	1021	2 I39207	leukocyte surface
19	164.5	5.1	3707	2 S18252	heparan sulfate pr
20	160	5.0	739	2 J50675	vascular cell adhe
21	159.5	4.9	1052	2 B49120	fibroblast growth
22	151.5	4.7	1443	2 I50600	neogenin - chicken
23	151	4.7	378	2 PH1379	vascular cell adhe
24	148.5	4.6	1239	2 A32579	neuroglian - fruit
25	147.5	4.6	799	2 S18209	fibroblast growth
26	147.5	4.6	1021	2 A37112	contactin precursor
27	147.5	4.6	898	2 A40114	fasciclin II precu
28	147	4.6	702	2 A36319	carcinoembryonic a
29	146.5	4.5	1088	2 IUXLNL	neural cell adhesi
30	144	4.5	538	1 JC2457	vascular cell adhe
31	143	4.4	1535	2 S46224	peroxidasin - frui
32	143	4.4	620	2 JH0593	Schwann cell myeli
33	142	4.4	2629	1 I38344	titin, cardiac mus
34	142	4.4	739	2 JQ0581	vascular cell adhe
35	141	4.4	1018	2 JQ4211	neural adhesion pr
36	141	4.4	267	2 A38442	probable tumor sup
37	140.5	4.4	725	2 J50099	neural cell adhesi
38	140	4.3	1091	1 I3CHNL	neural cell adhesi
39	140	4.3	1010	2 JU0094	Flil protein precur

ALIGNMENTS

RESULT 1
ORHUGS
secretory component precursor - human
N:Alternate names: poly-Ig receptor; polymeric immunoglobulin receptor
N:Contains: free secretory component; transmembrane secretory component
C:Species: Homo sapiens (man)
C>Date: 28-Aug-1985 #sequence.revision 23-Aug-1996 #text-change 13-Mar-1998
C:Accession: A46537; A55284; I38115; A32263; S38978; S13453; A02112
R:Krajci, P.; Kvale, D.; Tasken, K.; Brandtzaeg, P.
Eur. J. Immunol. 22, 2309-2315, 1992
A:Title: Molecular cloning and exon-intron mapping of the gene encoding human transme
A:Reference number: A46537; MUID:92387236
A:Accession: A46537
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-764 <KRA>
A:Cross-references: GB:S43449; NID:9255097; PID:9255098
A:Experimental source: leukocytes
A:Note: sequence extracted from NCBI backbone (NCBIP:113253)
R:Krajci, P.; Grzeschik, K.H.; Geurts van Kessel, A.H.; Olaisen, B.; Brandtzaeg, P.
Hum. Genet. 87, 642-648, 1991
A:Title: The human transmembrane secretory component (poly-Ig receptor): molecular cl
A:Reference number: A55284; MUID:92039621
A:Accession: A55284
A:Molecule type: mRNA
A:Residues: 1-764 <KR2>
A:Cross-references: GB:S62403; NID:9238235; PID:9238236
A:Experimental source: colonic adenocarcinoma cell line
A:Note: sequence extracted from NCBI backbone (NCBIN:62403, NCBIP:62408)
R:Piskurich, J.F.; France, J.A.; Tamer, C.M.; Willmer, C.A.; Kaetzel, C.S.; Kaetzel,
Mol. Immunol. 30, 413-421, 1993
A:Title: Interferon-gamma induces polymeric immunoglobulin receptor mRNA in human int
A:Reference number: I38115; MUID:93205018
A:Accession: I38115
A:Molecule type: mRNA
A:Residues: 1-764 <RES>
A:Cross-references: EMBL:X73079; NID:9456345; PID:9456346
A:Note: submitted to the EMBL/GenBank/DBJ databases by J.F. Piskurich, February 1994
R:Krajci, P.; Solberg, R.; Sandberg, M.; Oyen, O.; Jahnsen, T.; Brandtzaeg, P.
Biochem. Biophys. Res. Commun. 158, 783-789, 1989
A:Title: Molecular cloning of the human transmembrane secretory component (poly-Ig re
A:Reference number: A32263; MUID:89149795
A:Accession: A32263
A:Molecule type: mRNA
A:Residues: 72-764 <KR3>
A:Cross-references: GB:M24559; NID:9514365; PID:9514366
R:Fallgreen-Gebauer, E.; Gebauer, W.; Bastian, A.; Kratzin, H.D.; Eiffert, H.; Zimmer
Biol. Chem. Hopps-Seyler 374, 1023-1028, 1993
A:Title: The covalent linkage of secretory component to IgA. Structure of sIgA.
A:Reference number: S38978
A:Accession: S38978
A:Molecule type: protein
A:Residues: 478-488; 517-526; 543-545 <FAL>
A:Note: disulfide bonds for unbound and IgA-bound forms
R:Eiffert, H.; Quentin, E.; Wiederhold, M.; Hillemeir, S.; Decker, J.; Weber, M.; Hil
Biol. Chem. Hopps-Seyler 372, 119-128, 1991
A:Title: Determination of the molecular structure of the human free secretory compone
A:Reference number: S13453; MUID:91315750
A:Accession: S13453
A:Molecule type: protein
A:Residues: 19-135; 'O', 137-157; 'D', 159-207; 'DE', 210-228; 230-233; 'N', 235-240; 'O', 242-2
R:Eiffert, H.; Quentin, E.; Decker, J.; Hillemeir, S.; Hufschmidt, M.; Klingmuller, D

contactin precursor
myelin-associated
myelin-associated
neuronal cell surf
fibrillin 1 precur
neural cell adhesi

Hoppe-Seyler's 2. Physiol. Chem. 365, 1489-1495, 1984
A:Title: The primary structure of the human free secretory component and the arrangement
A:Reference number: A02112; MUID:85128981
A:Accession: A02112
A:Molecule type: protein
A:Residues: 19-157, D', 159-207, 'DE', 210-228, 230-233, 'N', 235-240, 'Q', 242-261, 'Q', 263-279,
A:Note: paper in German with English abstract
C:Comment: As a 100k transmembrane receptor for polymeric immunoglobulins, secretory com-
ylated, forms interchain disulfide bonds, undergoes proteolysis and transcytosis. Free s
C:Genetics:
A:Gene: GDB:PIGR
A:Cross-references: GDB:120280; OMIM:173880
A:Map position: lq31-lq41
A:Introns: 15/1; 130/1; 349/1; 460/1; 569/1; 629/2; 670/1; 714/1; 733/3
A:Note: the first intron occurs before the initiator codon
C:Complex: monomeric as a transmembrane receptor or free in mucosal secretions; heterod-
amers; hetero-22-mer composed of one chain of secretory component, one chain of immunogl-
C:Superfamily: secretory component; immunoglobulin homology
C:Keywords: duplication; glycoprotein; immunoglobulin receptor; phosphoprotein; transcyt-
F:1-18/Domain: signal sequence #status predicted <SIG>
F:19-764/Product: transmembrane secretory component #status predicted <MATW>
F:19-577/Product: free secretory component #status experimental <MATP>
F:33-112/Domain: immunoglobulin homology <IGV1>
F:145-222/Domain: immunoglobulin homology <IGV2>
F:250-327/Domain: immunoglobulin homology <IGV3>
F:364-443/Domain: immunoglobulin homology <IGV4>
F:475-546/Domain: immunoglobulin homology <IGV5>
F:639-661/Domain: transmembrane #status predicted <TMW>
F:662-764/Domain: intracellular #status predicted <INT>
F:40-110, 56-64, 152-220, 257-325, 271-279, 371-441, 395-482, 544-503/Disulfide bonds:
F:183, 90, 135, 186, 421, 469, 499/Binding site: carbohydrate (Asn) (covalent) #status experime
F:486-520/Disulfide bonds: (in Ig-unbound form) #status experimental
F:486/Disulfide bonds: interchain (to IgA chain-192) #status experimental
F:520/Binding site: cysteine (Cys) (covalent) (in Ig-bound form) #status experimental
F:577-578/Cleavage site: Lys-Ala (unidentified proteinase) #status experimental
F:673/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 100.0%; Score 3226; DB 1; Length 764;
Best Local Similarity 99.8%; Pred. No. 4.3e-223;
Matches 607; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MLFLVLTCLLAVFPALSTKSPIFGPEEVSVEGNSVITCYPPPTSVNRHTRKYWCROGA 60
Db 1 MLFLVLTCLLAVFPALSTKSPIFGPEEVSVEGNSVITCYPPPTSVNRHTRKYWCROGA 60

Qy 61 RGCITLISSEGVSSKYAGRANLTNFPNGTFFVNIQALSDODSGRYKCGIGNSRGLS 120
Db 61 RGCITLISSEGVSSKYAGRANLTNFPNGTFFVNIQALSDODSGRYKCGIGNSRGLS 120

Qy 121 FDVLSVSGPGLNDTKYVTDLGRVTINCPFKTENAKRKSLYKQIGLYPVLVIDSS 180
Db 121 FDVLSVSGPGLNDTKYVTDLGRVTINCPFKTENAKRKSLYKQIGLYPVLVIDSS 180

Qy 181 GYNPNYTRIRLDIOGTGQLLFSVVIINQLRSLDAGQYLCOAGDDSNKKNADLOVLKP 240
Db 181 GYNPNYTRIRLDIOGTGQLLFSVVIINQLRSLDAGQYLCOAGDDSNKKNADLOVLKP 240

Qy 241 EPELYVEDLRGVTFFCALGPEVANVAKFLCRQSSGENDGVVNTLGKRAPAFEGRILLN 300
Db 241 EPELYVEDLRGVTFFCALGPEVANVAKFLCRQSSGENDGVVNTLGKRAPAFEGRILLN 300

Qy 301 PODKDGSSVITGLRKEDAGRYLCAHSDGQLQEGSPICQAWOLFVNEESTIPRSPVYVK 360
Db 301 PODKDGSSVITGLRKEDAGRYLCAHSDGQLQEGSPICQAWOLFVNEESTIPRSPVYVK 360

Qy 361 GVAGSSVAVLCPYNRKESIKYWCLEWGAQNGRCPLLDVDSGQWKAQYEGRLSLEEPG 420
Db 361 GVAGSSVAVLCPYNRKESIKYWCLEWGAQNGRCPLLDVDSGQWKAQYEGRLSLEEPG 420

Qy 421 NCTFTVILNQLSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTAVLGETLKV 480
Db 421 NCTFTVILNQLSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTAVLGETLKV 480

Qy 481 PCHFPCKFSSYEKYCKWKNNTGCOALPSQDEGSPKAFVNCDSNRLVSLTNLTVRADEG 540
Db 481 PCHFPCKFSSYEKYCKWKNNTGCOALPSQDEGSPKAFVNCDSNRLVSLTNLTVRADEG 540

Qy 541 WYCGVKGQHFYGETAAVYVAVERKKAAGSRDVSALAKADAAPDEKVLDSGFRFENKAIQ 600
Db 541 WYCGVKGQHFYGETAAVYVAVERKKAAGSRDVSALAKADAAPDEKVLDSGFRFENKAIQ 600

Qy 601 DPLRFAEZ 608
Db 601 DPLRFAEZ 608

RESULT 2
145956
polymeric immunoglobulin receptor - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 19-Dec-1997 #sequence_revision 19-Dec-1997 #text_change 09-Apr-1998
C:Accession: 145956
R:Kulseth, M.A.; Krajci, P.; Myklebost, O.; Rogne, S.
DNA Cell Biol. 14, 251-256, 1995
A:Reference number: 145956; MUID:95186063
A:Accession: 145956
A:Status: preliminary; translated from GB/EMBL/DBDJB
A:Molecule type: mRNA
A:Residues: 1-757 <KUL>
A:Cross-references: GB:I04797; NID:g388279; PID:g388280
C:Superfamily: secretory component; immunoglobulin homology
F:145-222/Domain: immunoglobulin homology <IMM1>

Query Match 67.5%; Score 2178; DB 2; Length 757;
Best Local Similarity 67.2%; Pred. No. 3.9e-148;
Matches 405; Conservative 80; Mismatches 114; Indels 4; Gaps 3;

Qy 6 LTCLLAVFPALSTKSPIFGPEEVSVEGNSVITCYPPPTSVNRHTRKYWCROGAGCCI 65
Db 6 LACLAIFPVVYSKSPIFGPEEVSVEGNSVITCYPPPTSVNRHTRKYWCROGAGGRC 65

Qy 66 TLISSEGVSSKYAGRANLTNFPNGTFFVNIQALSDODSGRYKCGIGNSRGLSFDVSL 125
Db 66 TLISSEGVSSKYAGRANLTNFPNGTFFVNIQALSDODSGRYKCGIGNSRGLSFDVSL 125

Qy 126 EVSQDPAQASHAHVYTDLGRVTINCPFKTENAKRKSLYKQIGLYPVLVIDSSGYVNP 185
Db 126 EVSQDPAQASHAHVYTDLGRVTINCPFKTENAKRKSLYKQIGLYPVLVIDSSGYVNP 185

Qy 186 NYTRIRLDIOGTGQLLFSVVIINQLRSLDAGQYLCOAGDDSNKKNADLOVLKPEPELV 245
Db 186 SYKDRAHISILGTLNLFVSVIINRVKLSADAGMYVCQAGDDAKADKINIDLOVLEPELV 245

Qy 246 YEDLRGVTFFCALGPEVANVAKFLCRQSSGENDGVVNTLGKRAPAFEGRILLNPOKD 305
Db 246 YGDLRSSVTFDCSLGPEVANVAKFLCRQSSGENDGVVNTLGKRAQDFQGRIVSPKD-N 304

Qy 306 GSFVSVITGLRKEDAGRYLCAHSDGQLQEGSPICQAWOLFVNEESTIPRSPVYVKVAGS 365
Db 306 GSFVSVITGLRKEDAGRYLCAHSDGQLQEGSPICQAWOLFVNEESTIPRSPVYVKVAGS 365

Qy 366 SVAVLCPYNRKESIKYWCLEWGAQNGRCPLLDVDSGQWKAQYEGRLSLEEPGNGTFT 425
Db 366 SVAVLCPYNRKESIKYWCLEWGAQNGRCPLLDVDSGQWKAQYEGRLSLEEPGNGTFT 425

Qy 426 VILNQLSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTAVLGETLKVPCHP 485
Db 426 VILNQLSRDAGFYWCLTNGDTLWRTTVEIKIIEGPNLKVPGNVTAVLGETLKVPCHP 485

Qy 486 CKFYSFEKYCKWKNNTGCOALPSQDEGSPKAFVNCDSNRLVSLTNLTVRADEGQWCG 545
Db 486 CKFYSFEKYCKWKNNTGCOALPSQDEGSPKAFVNCDSNRLVSLTNLTVRADEGQWCG 545

A:Description: The cloning, tissue specific expression and interspecies sequence comparison of the human *CD44* gene.
A:Reference number: S48841

Query Match 67.5%: Score 2177; DB 2; Length 757;

RESULT	QRRTGS	4
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

Query Match: 64.0%: score 2064.5: DB 1: Length 769:

Query Match	64.0%	Score 2064.5	DB 1	Length 789
Best Local Similarity	62.6%	Pred. No. 5.3e-140		
Matches	385	Mismatches	125	Indels 9
Conservative	96			Gaps 5

Qy	1	MLFLVTLCLLAVFPAISTKSP	IFGPEEVNSVEGNSVSI	ITCYPP	TSVNRHTRKRYWCROGA	60
Db	1	MLSLFALLVTVFGSVTSQSP	IFGPDQDVSSIEGNSVSI	ITCYPP	TSVNRHTRKRYWCROGA	60
Qy	61	RGCGITLISSEGYSSKYAGRAN	TNPENG	TFVWIA	IOSDDSGRYCKGLGINSRGLS	121
Db	61	NGVCATLISSNGYLSKEYSGR	ASLINPEN	STFVWIA	IAHLEOTEDTGSYCKGLGTNRGLF	121
Qy	121	FDVSLVSVGPGCLLNDT	KVYTV	DGLGRV	TTCNCFKTEKNAQRKSLYKQIGLPLVPLVIDSS	181
Db	121	FDVSLVSVGPFPNDT	HVYKDI	GRVTV	TECFKSGNAHSSKSLCKRGEACEVVIDST	181

```

F:46-115,155-225,260-324,369-438,478-538/Disulfide bonds: #status experimental
F:108/Binding site: carbohydrate (Asn) (covalent) #status experimental
F:418/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match          44.5%;   Score 1435.5;   DB 1;   Length 773;
Best Local Similarity 49.7%;   Pred. No. 5.2e-95;
Matches 316;   Conservative 87;   Mismatches 186;   Indels 47;   Gaps 18;

Qy  1  MLFVFLTCLLAVPAISTK-----SPIFPEEVNSVEGNSVITCYYPFTSVNRHTRKY 54
      |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db  1  MALFLLTCLLAVSAAPQAQSSLLGCPSSIFGPGEVNVLGDSVSTCYPTTSTRHSRKF 60
      |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

Qy  55  WCROGARGGCTILISSGGYSSKVGAGRANTNPENCTFVVNTAQLSDSDSGRYKCGLGI 114
      |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db  61  WCREESGRCTIL-ASTGYTSQEFSGRKKLTDPDGKEFVTVVDQLTQNDSSGYKCGGV 119
      |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

Qy  115  NSRGLSFDVLSVYSGQGLLNDTRKVYTVDLGRVITINCPFKTENAAQRKSLYKOIGLYPV 174
      |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db  120  NGRGLDFGVNVLVSQKPE--PDDVVYKQYESYTVTITCPTYATRLQKKSFYKVEDGELV 177
      |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

Qy  175  LVIDSQG--YVNPNTGRIRLDTQGTQQLLFSVVIINOLRLSDAGQYLCQAGDDSDNSNKK 232
      |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db  178  LIIDSSSKEARDPYKGRITLQGSTTAKETFTYIKHLQNDAGQYVCQSGSDPTAEQON 237
      |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

Qy  233  ADLQVLKPEELVYEDLURGSYTFHCALGPEVANVAKFLCRQSGGENDCVVNTLGKRAPA 292
      |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||
Db  238  VDLRLL--TFGLLYGNLGGSVTFECALDSEDANAVASL-RQVRGN--VVIDSOGTIDPA 292
      |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

Qy  293  FEGRIILLNPQDGFSFVITGLRKEDAGRYLCGAHSDGOLQEGSPIOAWOLFVNEESTI 352
      |||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||:|||

```

```

QY 353 PRSPVTVGVAGSSVAVLCYPYNRKESIKYKWLWEAGQNGRCPLLVDS-EGWYKAQYEG 411
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 351 SRSPVLVKGPGGSGVTIRCPYKPRSDSHLQLYLWEGSQTRH--LLVDSGEGLVKQDYTG 408
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 412 RLSLEEPGGNGTFTVILNQLTSRDAGFYWCLTNGDTLWRTTVEIKIIEGEPNLKVPNGVT 471
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 409 RLALFEPGGNGTFSVLNQLTAEDGEGFYWCSDDESLLTSVKLQIWDGEPSPTI-DKFT 467
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 472 AVLGETLKVPCHPFCKPCKSFSEYKWKNNWTCGALPQSDGEPGSKAFYNCDSNRLVLT 531
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 468 AVQGEPEVITCHPCKYFSEKWKYKNDHGCDLPTK-LSSSGDLVKCN-NNLVLTLTL 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 532 NLVTRADEGWYCWGVKOGHYGETAAV-----YVAVERKAAGSRDVSIAKADAAPD 583
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 526 DSVSDEDEGWYCWGAKDGHFEFEVAAVRVELTEPAKVAVEPAKV----PVDPAKAAAPA 581
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 584 EKVLDS-----GFREIENK----AIQDPLFAEZ 608
    : : : : : : : : : : : : : : : : : : : : : :
Db 582 EERAKARCPVRRRQWYPLSRKLTSCPEPRLAAE 617
    : : : : : : : : : : : : : : : : : : : : : :
RESULT 6
PN0568
Connectin 3B - chicken (fragment)
N:Alternate names: Cn3B protein
C:Species: Gallus gallus (chicken)
C:Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 17-Mar-1999
C:Accession: PN0568
R:Maruyama, K.; Endo, T.; Kume, H.; Kawamura, Y.; Kanzawa, N.; Nakauchi, Y.; Kimura,
  Biochem. Biophys. Res. Commun. 194, 1288-1291, 1993
A:Title: A novel domain sequence of connectin localized at the 1 band of skeletal muscle
A:Reference number: PN0568; MUID:93356802
A:Accession: PN0568
A:Molecule type: mRNA
A:Residues: 1-1323 <MAR>
A:Cross-references: DDBJ:D16541; NID:g931629; PID:d1004495; PID:g931630
A:Experimental source: skeletal muscle
C:Comment: This protein string-like single molecule spans from the 2 line to the M 11

```

```
Query Match          6.5%; Score 210.5; DB 2; Length 1323;  
Best Local Similarity 21.9%; Pred. No. 4.6e-07;  
Matches 149; Conservative 95; Mismatches 240; Indels 195; Gaps 35;
```

QY 6 LTCLLAVFPAISTKSPIGPEVN--SYE-GNSVSTICYI---PPTSVNRHTRKKWCROG 59
Db :|: |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
75 ISCTHLF-----VKEPAFVMKNVDLSEVGKANLILECTYTGTPISVT-----WKNG 124
QY 60 ARGGCITLISSEGYYSSYAGRANLTNFPENGTFVVNIQAQLSQDDSGRYKCGLGINSRG 119
Db :|: |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
125 -----VILKHSEKCSITTETSA--ILLEIPNSKLEDOGOYSCHIENDSGOD 168
QY 120 SFDVSLEYSQGLLNDRKVITYDVLGRVTINCPKTENAQRKSLSYKQIGLYPLVLIDS 179
Db ::::: |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
169 NCHGAITILEPPYFTLPLEPVQTVGVDSASLOC-----OVAGTPEMIVSW 213
QY 180 SGYNPNNTGRIRLDIQGTGOLLF-----SVVINOLRLSDAGOYLCO----AGDDSNNK 230
Db :|: |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
214 -----YKGTKLRGTATVMKHFKNQVATLVFSQDDSDGEYICKVENTVEATSSL 266
QY 231 KNADLOVLKP-----EPELVIEDLRGSVFTHCAL-GPE-----263
Db :|: |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
267 LTQERKLPPSFTRKLRDVHTVGLPVTFDCGIAGESEPTEYSWFKDNVRKYEDYNVHTSF 326
QY 263 VANTA-----KFLCRQS-----SGENCDEVNTLGKRAPAFEGRILLNPQD-- 304
Db :|: |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
327 IDNVAILQLKTDKSLMGQYTCTASNAGTASSSKGLVLTGKTPPFDTPT--TPVDGI 384
QY 304 -----KD-----GSFSV-----VITGLRKEDADGYLC 325
Db ||||: |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
385 IGESADFECHISGTQPIRVTWAKNQEI RTGGNYOISIVENTAHTLIRVDRGDSGKTC 444
QY 326 GAHSDGDLQEGSPIQAWOLFVNEESTIPRSPTVKGVAAGSVAYLCLPYNRKESKSIKYMC 385
Db :|: |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
445 YASN----EVCKSDCSTAQLNVKERKT---PPTFRKL---SEAV---EETEGNELKL-- 489
QY 386 LWEGAQNRCRP LLVDSEGWVAAEQEGRILLSLEEPG--NGFTVILNQLTSRDAGFYWLCT 443
Db -EGRVAGSQPLTV---SWYKNNOEVHSHPCEITSFKNNTLLLHHIKSVQGSDAGLYTKCV 543
QY 444 N---GDTLWRRTWTBEIKIEGEPNLKVPNGNAV-LGETLKVPCHPEPKCFSSYEKYMKWN 499
Db :|: |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
544 SNEAGSVLCTS SVVIREPKPPVDPQLQPAATEBGDILQLSCH---VRGSEPIRIOWL 599
QY 500 NTGCCALPSQDEGSPSKAFVNDENSRLVSLTNLVTRADEGWYCGVYQGHEFYG-ETAAY 558
Db :|: |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
600 KAGREIRAS--ERGCSFSPAN-----GVALLLEAATAKSDSGEYVC--KASNVAGTDTCSR 650
QY 559 YVAVEERKAAGSRDVSIAK 577
Db :|: |: | |: | |: | |: | |: | |: | |: | |: | |: | |: |
651 KVTVKEKNAAL-----VSAAK 665

RESULT 7
I38346
elastic titin - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 05-Jun-1998
C:Accession: I38346
R:Labeit, S.; Kolmerer, B.
A:Title: Titins: giant proteins in charge of muscle ultrastructure and elasticity.
A:Reference number: A57430; MUID:96026330
A:Accession: I38346
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-7962 <RES>
A:Cross-references: EMBL:X90569; NID:g1017426; PID:g1017427
C:Genetics:
A:Gene: GTN
A:Cross-references: GDB:l27867; OMIM:188840
A:Map position: 2q31-2q31

Query Match 6.18; Score 197; DB 2; Length 7962;
Best Local Similarity 19.9%; Pred. No. 5e-05;
Matches 137; Conservative 91; Mismatches 245; Indels 216; Gaps 31;

QY 10 LAVPFAISTSPIGPEVANSVEGNSYITCIYP--PTSNNRHTRYKWCQGARGCII 66
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 1336 LIIPSTTKL-----KKMDSIKGSFDLECTIVAGSHPIIQ-----WFKD----- 1377
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 67 LISSEGVSSYAGRANLTFFPENGTVFVAIAQSODSGRYKCGLGINSRGLSFYDSLE 126
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 1377 --DQEISASEKYK----FSFHDTAF--LEISOLEGDSGTYTCS--ATNKAG----- 1420
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 127 VSQPGLLNDTKYVDLGRVTINCSPFTENAKRKSLYKQLGLPYVLVIDSSGYNPN 186
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 1420 -----HNQCSGHILTKEPPYFEVKPSQD-----YNPN 1447
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 187 YTGRIRLDIQGTGOLLF-----SVVINQLRLSDAQYLCOA 222
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 1448 TRVOLKALVGTAPMWTIKWFKNDELHSGAARSVMKDDTSTSELFAAKATDSYTIQOL 1507
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 223 GDDSNSNKKNIADLOV-----LKPEPELYVEDLRG--SVTFHCAL--GPEVANVAKFLCR 272
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 1508 SNDVGTATSKATLFVKEPPOFIKKPSPLV--LRNGOSTTFECQITGTPKIRVSWYL-- 1563
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 273 QSSGENCDVVNTIGKRAPAFEGRIILNPDKOGSFSVITGURKEDAGRYLCAHS DGQ 332
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 1563 ---DGNE---ITAQKHGISF-----IDGLATFOISGARVENSGTYVCEARNDA- 1606
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 333 LOESGPQAQOLFVNRESTIPRSPTVVKGAGSSVALC-----PV-----NRKESKS 380
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 1606 ---GTASCSIELKVKEPTIRELKPVEVVKYSDOVELECEVTGTPPEVTWLKNNREIRS 1662
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 381 IKYWCL-----WEGAONGRCPLLVDSEGWMKAQYEGRLSLEBEPG-----NG 422
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 1663 SKKYTLDRVSFNLHITKCDPSDTGEYQIVSNEG--SCSCSTRVALKEPFPFIKKIEN 1721
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 423 TFTVILNQLT-----SRDAGFYW-----CLNGDPDLMT-----TVEIKIEGPN- 464
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 1722 TTTVLKSSATFQSTVAGSPPISTITLWKDDOILDDEDNVIYSFVDSVATLQIRSYDNGHSG 1781
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 464 -----LKPQGNV---TAVIGETLKVPCHPCKPFSSYEKWKCKWN 500
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 1782 RYCQANKESGVERCYAFLIVQPAQIVEKAKSVDVTEKDPMTLECVAAGTPELVKVKALK 1841
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 501 TGCOALPSQDEGPSKAFVNCDENSRLVSLTNLVTRADEGWYCWGKYGQHFGETAAYV 560
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 1842 DGKIQVPSP-----YFSMSFENNVAFSIQSYMKQDSGOYTFKV-ENDFGSSCCDAYL 1893
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 561 AVERKKAAGSRDYSLAKADAAPDEKVLDS 589
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
DB 1894 RVLDONTPPSFTKKLTAKMD-----KVLGS 1917
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

RESULT 8
I37243
CMRF-35 antigen - human
C:Species: Homo sapiens (man)
C>Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 28-Feb-1997
C:Accession: I37243
R:Jackson, D.G.; Hart, D.N.; Starling, G.; Bell, J.I.
Eur J. Immunol. 22, 1157-1163, 1992
A>Title: Molecular cloning of a novel member of the immunoglobulin gene superfamily h
A:Reference number: I37243; MUID:92249405
A:Accession: I37243
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 1-224 <RES>
A:Cross-references: EMBL:X66171; NID:g9396169; PID:g9396170
C:Genetics:
A:Gene: CMRF35

Query Match 5.8%; Score 187; DB 2; Length 224;
Best Local Similarity 29.7%; Pred. No. 1.9e-06;
Matches 51; Conservative 20; Mismatches 75; Indels 26; Gaps 5;

QY 356 PTVVKGAGSVAVLCPYRNKESKSIKYLWLGAGNCRPLLVDSGVMVKAQYEGRLSL 415
DB 28 PMTVAGVGSLSVQCRYEKEHRLNFKWC--RPPILCKDKIVETKG-SAGKRNGRVSI 84
QY 416 LEEPGNGTFTVILNOLTSRDAGFYWGLTNGDTLWRTTVEIKIEGPNLKPVGNTVA--- 473
DB 85 RDSPANLSFTVTLLENTEEDAGYWC--GVDTPWLRFHDPIVEVEVSVPAGTITASSP 142
QY 473 -----VLGETLKVPCHPKCFSSYKWKNNWTCQALPSQDEGSPKAFVN 519
DB 143 QSSMGTSGPPTKLPVH-----TWPSVTRKDSPEPSPHGSLFSN 181

RESULT 9
BIG-1 protein - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 18-Sep-1998
C:Accession: I58164
R:Koshihara, Y.; Kawaaki, M.; Tani, A.; Tamada, A.; Negata, S.; Kagamiyama, H.; Mori, K.
Neuron 13, 415-426, 1994
A:Title: BIG-1: a new TAG-1/F3-related member of the immunoglobulin superfamily with neu
A:Reference number: I58164; MUID:94338697
A:Accession: I58164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1028 <RES>
A:Cross-references: EMBL:U11031; NID:g563132; PID:g563133
C:Genetics:
A:Gene: BIG-1
C:Superfamily: contactin; fibronectin type III repeat homology; immunoglobulin homology

Query Match 5.7%; Score 185; DB 2; Length 1028;
Best Local Similarity 21.9%; Pred. No. 2.2e-05;
Matches 132; Conservative 91; Mismatches 243; Indels 136; Gaps 31;

QY 25 PREVNSVENSITCY---YPTSVNRTRKYWCQARGGCITLISSEGVSKYAGR 81
DB 234 PETLPAAGSTVKLECFALGNVPVQIN-----WRR-----SDG---WPFPFK 272
QY 82 ANLTNPENGTFVNTAQLSQDSGRYKGLGINSRGLSFDVSLEVSQGPGLNDTKVYT 141
DB 273 IKLRK--NG--VLETPNQOEDTGSYEC-IAENSRGN-----VARG-RLTYAKPKW 320
QY 142 V----DLGRVTINCPFTENAQRKSLYKQIGLYPVLVIDSSGVYNPYTGRIRLDIQ 197
DB 321 VOLLKDVETAVEDSLYWECRASGKPKPSYRWLKNGDALVLE-----RIQIEN 368
QY 198 TQOLFSVVINOLRLSDAQYLCOAGDDSNKKNADLOVLPPELYVEDLRGSVTHC 257
DB 369 G-----ALTIANLVSDSGMFOCIAENKHLIYSSAELKVLASAPDFSNNPKMKIQV-- 422
QY 258 ALGPEVANNVAKELCROSSGNC-----DVVNTLGRKPAFEGRI--LLNPQDDGSGFS 309
DB 422 ----QVGSVLILDKRPSAPRSLFSFKKGDVVRE-----QARISLN---DGGLK 465
QY 310 VVITGLRKEDAGRYLGGASDGOQLBGSPIQAWOLFVNEESTIPRPTVVKGAVSSVAV 369
DB 466 IM--NVTKADAGIYTCIAEN---QFGKANGTQLVVTETRIILASPMNDVAVGESIIL 519
QY 370 LCPYRNKESKSIKYLWLGAGNCRPILVDSGVMVKAQYEGRLSLLEP-----CN 421
DB 520 PC-----QVQHPDLIDIMFAN---YFNGTLTDFKDKGSHFEKVGGS 557
QY 422 GTFTVILNOLTSRDAGFYWCL--TNGDTLWRTTVEIKIEGEP-----NLKVP--GNVTAV 473
DB 558 SSGDLMIKRLQKHSKYYCMVQTGVDSV--SSNAEL-IVRSGPGPENKVDIETDTTAQ 615

QY 474 LGETLKVPCHPPC---KFSSEYKWKNNWTCQALPSQDEGSPKAFVNCDENSRLVSLT 530
DB 616 LSWTEGTDHSPVISYAVQARTPFSGVGNV--RTVPEAIDGKTRTATVWELNP--WVEYE 672
QY 531 LNLVTRADEGWYKVGKOGHEFYGTAAVYVAVEB--RKAAGSRDYSLAKADAAPKEKVLDS 589
DB 673 FRVVASNKIGGSEPLSPSEKIVRTEEAPEVAPSEVSGGGGSRSELVITWDPVPEELQNG 732
QY 590 GF 591
DB 733 GF 734

RESULT 10
A38096
perlecan precursor - human
N:Alternate names: basement membrane heparan sulfate proteoglycan; heparan sulfate pr
C:Species: Homo sapiens (man)
C:Date: 07-Apr-1994 #sequence_revision 07-Apr-1994 #text_change 15-Jan-1999
C:Accession: A38096; S19256; S77946; A41059; A40306; B33625; A41736
R:Murdoch, A.D.; Dodge, G.R.; Cohen, I.; Tuan, R.S.; Iozzo, R.V.
J. Biol. Chem. 267, 8544-8557, 1992
A:Title: Primary structure of the human heparan sulfate proteoglycan from basement me
tor, laminin, neural cell adhesion molecules, and epidermal growth factor.
A:Reference number: A38096; MUID:92235084
A:Accession: A38096
A:Molecule type: mRNA
A:Residues: 1-4391 <MUR>
A:Cross-references: GB:M85289; NID:g184426; PID:g184427
R:Kallunki, P.; Tryggvason, K.
J. Cell Biol. 116, 559-571, 1992
A:Title: Human basement membrane heparan sulfate proteoglycan core protein: a 467-kD
ell adhesion molecules, and epidermal growth factor.
A:Reference number: A41736; MUID:92112994
A:Accession: S19256
A:Molecule type: mRNA
A:Residues: 1-57, 'D', '59-434, 'A', 436, 'FL', 438-449, 'Q', 451-502, 'A', 503-792, 'K', 794-908,
71-2979, 'H', 2981-2994, 'G', 2996-3167, 'T', 3169-3240, 'R', 3242-3426, 'R', 3428-3631, 'Q', 363
A:Cross-references: EMBL:X62515; NID:g29469; PID:g29470
R:Kallunki, P.; Eddy, R.L.; Byers, M.G.; Kestilae, M.; Shows, T.B.; Tryggvason, K.
Genomics 11, 389-396, 1991
A:Title: Cloning of human heparan sulfate proteoglycan core protein, assignment of th
A:Reference number: A41059; MUID:92120660
A:Accession: A41059
A:Molecule type: mRNA
A:Residues: 'RT', 892-908, 'R', 910-1101, 'L', 1103-1132, 'L', 1134-1221, 'L', 1223-1397 <KA2>
A:Cross-references: GB:S76436; NID:g243370; PID:g243371
R:Dodge, G.R.; Kovalszky, I.; Chu, M.L.; Hassell, J.R.; McBride, O.W.; Yi, H.F.; Iozzo
Genomics 10, 673-680, 1991
A:Title: Heparan sulfate proteoglycan of human colon: partial molecular cloning, cell
A:Reference number: A40306; MUID:91365376
A:Accession: A40306
A:Molecule type: mRNA
A:Residues: 1018-1409, 'G', 1407-1409, 'G', 1411-1465 <DOB>
A:Cross-references: GB:M64283; NID:g184424; PID:g184425
R:Heremans, A.; van der Schueren, B.; De Cock, B.; Paulsson, M.; Cassiman, J.J.; van
J. Cell Biol. 109, 3199-3211, 1989
A:Title: Matrix-associated heparan sulfate proteoglycan: core protein-specific monoc
anes.
A:Reference number: A33625; MUID:90078352
A:Accession: B33625
A:Molecule type: protein
A:Residues: 1379-1384, 'X', 1386-1388, 'X', 1390-1398 <HE2>
A:Accession: A33625
A:Molecule type: protein

A:Residues: 2166-2171,'X',2173-2175,'X',2177-2185<HE3>
A:Note: peptide potentially matches four different regions of sequence shown
C:Genetics:
A:Gene: GDB:HSPG2
A:Cross-references: GDB:126372; OMIM:142461
A:Map position: lp36.l-1p36.1
C:Superfamily: LDL receptor ligand-binding repeat homology; laminin G repeat homology; laminin G repeat homology; transmembrane
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein; heparan sulfate;
F:1-21/Domain: signal sequence #status predicted <SIG>
F:23-4391/Product: perlecan #status predicted <MAT>
F:22-193/Domain: I <DOM1>
F:194-530/Domain: II <DOM2>
F:199-234/Domain: LDL receptor ligand-binding repeat homology <LDL1>
F:285-319/Domain: LDL receptor ligand-binding repeat homology <LDL2>
F:325-359/Domain: LDL receptor ligand-binding repeat homology <LDL3>
F:368-403/Domain: LDL receptor ligand-binding repeat homology <LDL4>
F:531-1676/Domain: III <DOM3>
F:1563-1610/Domain: laminin-type EGF-like homology <EG7>
F:1677-3686/Domain: IV <DOM4>
F:2007-2034/Domain: transmembrane #status predicted <TRM>
F:3687-4391/Domain: V <DOM5>
F:3953-4106/Domain: Laminin G repeat homology <LG2>
F:4149-4151/Region: motor neuron attachment (L-R-E) motif
F:4299-4301/Region: motor neuron attachment (L-R-E) motif
F:65.71.76/Binding site: heparan sulfate (Ser) (covalent) #status predicted
F:89.554,1755,2121,3072,3105,3279,3780,3936,4068/Binding site: carbohydrate (Asn) (covalent) #status predicted
F:2295,3933,4179/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted

```

Query Match          5.7%; Score 182.5; DB 2; Length 4391;
Best Local Similarity 21.1%; Pred. No. 0.00024;
Matches 136; Conservative 86; Mismatches 229; Indels 195; Gaps 28;

Qy 13 FPAJSTKSPFEGPEVNSVEGNSVITCYYPPTSVNRHTRKYCQARGGCITLISSEG 72
Db 2335 YPAGSTQPIRETPASSQVAEGQTLNLCVPGQS---HAQVTWHKRG-----G 2379

Qy 73 YVSKYAGRANLTNFPENGTFVFNIAQLSQDDSGRYKCGLGINSRGLSFDVLSLEVS----- 129
Db 2380 SLPRVHQTHGSL-----LRLYQASPADSGEYVC-----RVLGSSVPLEASVLVT 2423

Qy 129 -----QGPELLNDTKVYVVDLGRVTINCFFKENAQKRKSLYKQIGLYPLVLI 177
Db 2424 IEPAGSVPALGVPTPVRIESSSSQVAEGQTLNLC- LVAGQAHQAQVTHWKRGGSLPA--- 2480

Qy 178 DSSGYVNPNTGRIRLDIQGTGQLLFSVVINQLRLSDAGQYLCOAGDDSNSNKKNADLQV 237
Db 2480 -----RHQVHGSRLRL-----QVTPADSGEYVCRV--VGSSGTQEASVLV 2518

Qy 238 LKPEPELYVEDLRGS-----VTFHCALGPEVANVAKFLCRQSSGENCDVYN-----TLGK 288
Db 2519 -----TIQORLSGSHSGVAYPVRIESSASLAN---GHTLDNLCLVSAQPHITWYK 2569

Qy 289 RAFAFEGRILLNPQDKGSFSVVIITGLR-----KEDAGRYLCGAHSDGQLEGSPIOAWQ 343
Db 2570 RGGSLPSR-----HQIVGSRLEIPQVTPADSGEYVCHVSNAGSRETSLIVTIQ 2618

Qy 344 LFNVEESTIP-----RSPTVYVK-----VAGSSVAVLCPYNRKESKSIKYWC 385
Db 2619 --GGSSSHVSPPIRIESSPTVVEGQTLNLCVWARQPOAITWYK----- 2667

Qy 386 LWEGAQNCRCLPLVDSEGVWKAQYEGRLSILLEPENGCTFTVILNQLTSRDAGFYWCLTNG 445
Db 2667 -----GGSPLSRHQTHGSHLR-----LHQMVSADSGEYVCRRAN 2700

Qy 446 --DILWRTTVEIKIIEGEPNLKVPQNTAVL-----GETLKVCPCHPCPFSSSYEK 493
Db 2701 NIDAL-EASIVISVSPSAGSAPAGSSWPRIESSSHVHAEGEITLDLCNVPGQAHQA-- 2758

Qy 494 YWCKWNNTGCOALPQSDQEGPSKAFVNCNDSNRLVSLTLNLVTRADEGYWCVKQGHFYG 553
Db 2758 --VTWHKRG--GSLPSHHO-----TRGSRLRLHHVSPADSGEYVCRV--MGSSGP 2801

```

```

QY 554 ETRAAVVAVEERKAAGSRDVSIAKADAAPDEKVLDSGFREIENKAI 599
      I : I : I : I : I : I : I : I : I : I : I : I : I : I : I :
Db 2802 LEASVLVTIE---AGSSAVHVPAFGAPPRIIEFPSSSRVAEGOTL 2844

RESULT 11
I56171
B-cell adhesion protein CD22 - human
C:Species: Homo sapiens (man)
C:Date: 06-Sep-1996 #sequence_revision 06-Sep-1996 #text_change 29-Aug-1997
C:Accession: I56171
C:Wilson, G.L.; Naifeld, V.; Kozlov, E.; Menniger, J.; Ward, D.; Kehrl, J.H.
J. Immunol. 150, 5013-5024, 1993
A:Title: Genomic structure and chromosomal mapping of the human CD22 gene.
A:Reference number: I56171; MUID:93267103
A:Accession: I56171
A:Status: preliminary; translated from GB/EMBL/DBDJ
A:Molecule type: DNA
A:Residues: 1-664 <RES>
A:Cross-references: GB:S61375; NID:g385980
C:Genetics:
A:Gene: GDB:CD22
A:Cross-references: GDB:I27545; OMIM:107266
A:Map position: 19q13.1-19q13.1
A:Introns: 18/1; 120/1; 209/1; 297/1; 383/1; 471/1; 561/1; 593/2; 621/3

Query Match          5.4%; Score 173.5; DB 2; Length 664;
Best Local Similarity 22.5%; Pred. No. 7.9e-05;
Matches 145; Conservative 75; Mismatches 227; Indels 197; Gaps

```

Query Match	5.4%	Score 173.5	DB 2	Length 664
Best Local Similarity	22.5%	Pred. No. 7.9e-05		
Matches 145	Conservative	75	Mismatches 227	Indels 197
Gaps				
QY	25	PEBVSNEGVSITCYPTSVNRHTRKYWCROGAGGCITLISSEGYVSSKYAGRANL	84	
DB	28	PPRIQ--ESQEVTLICLLNFSCYGYPIQLWLEGVPMQAAVTSITIKSVPT-RSEL	84	
QY	85	TNPE---NGTFVFNIAQSDSGRYKCGLGINSRGLSFD-VLSVQSGPGLLNDTKVY	140	
DB	85	KFSPQWSHGKIYT--COL-QDADGKF-----LSNDTVQLNVKHTPKL--EIKVT	129	
QY	141	TVDL----GRTVINCFFKTENAKRKSLYKQIGLYPLVIVDSSGYNPNTVGRIRLDI	196	
DB	130	PSDAIVREGDSVTMTCEVSS-----TNPEYTVSWLK-D	162	
QY	197	GTG---QLLFSVVINRLSDAGQYLCOAGDSDNSNK-KNADLQV-LKPEPELYV---	247	
DB	163	GTSLKQNTFTLNLREVTKDQSGKCCQVSDNDVGPGRSEEVFLQVQVAPESVQILHSP	222	
QY	248	DLRGS-VTFHC-----ALGPEVANVAKFLCROSSGENSCDVVVNT	285	
DB	223	AVEGSOVEFLCNSLANPLTNTWTYHNGKEMQGRTEEVKHPKILPHAGTYSC-VAENI	281	
QY	286	L--GKRAPAFEGRIILNPQDKSGSVVITG--LRKEDAGRYLCGHAHSDG--QLQEGS	337	
DB	282	LGTQGRGPAG---LDVQYPKKVTVIQNPMPIREGDTVTLSGNTSSNPSTVRIEWK	337	
QY	338	PIQAWQ-----LFVN-----EESTI-----PRSPTV--VRGV-	363	
DB	338	PHGAWEPSLGLVIKQVGDWNTTITACARCSWCSWASPVALNVQYAPRDVVRVKIKPLS	397	
QY	363	---AGSSVAVLCPYNRKESKSYKWCLEMGAGNQRCPLLVDSEGWVRAQYEGRLSLLEEP	419	
DB	398	EIHSGNSVSLQDCDFSSHPKEVOFF--WE--KNGR--LL-----GKESQLN--	438	
QY	420	GNGTFVILNQLTSRDAGYFWCLTN--GDILWRT-TVEIKIIEGEPNLKV---PGNVTA	472	
DB	438	-----FDSISPEDAGYSICWVANSIGOTASKAWTLLEVYADAPRLRLRSMSPGD-QV	488	
QY	473	VLGETLVCPCHRPKFFSSYKVKWCKWNTTGOALPQSDQEGPSKAFVNCDENSRLVSLTN	532	
DB	489	MEGKSATLRCESDANDPVSHYTWFDWN---QSLDYHSQ-----KLRL	529	
QY	533	LVTRADEGWYCW---GVRQGHGFYETAAYVAVV---ERKAAG	569	

Db 530 PVKQHSYGAYWCOTNSVGRSPLSTLTLYVYSPETIGRVAVG 573

RESULT 12

I51669

tumor suppressor - African clawed frog

C:Species: Xenopus laevis (African clawed frog)

C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 28-Feb-1997

C:Accession: I51669

R:Piechall, W.E.; Reale, M.A.; Candia, A.F.; Wright, C.V.; Cho, K.R.; Fearon, E.R.

Dev. Biol. 166, 654-665, 1994

A:Title: Expression of a homologue of the deleted in colorectal cancer (DCC) gene in the

A:Reference number: I51668; MUID:95113183

A:Accession: I51669

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1427 <PIE>

A:Cross-references: EMBL:U10986; NID:g606873; PID:g606874

C:Genetics:

A:Gene: XDCCA

Query Match 5.3%; Score 170; DB 2; Length 1427;
Best Local Similarity 21.2%; Pred. No. 0.0004;
Matches 99; Conservative 76; Mismatches 189; Indels 104; Gaps 21;

QY 145 GRTVTINCPFKTENAQRKSL-----YKIGLXPVLVIDSSGVNPNYTGRIRLDIQGTQ 200

Db 54 GGNVLNC-----SAQSDGAPIKKWKGVLNLVIDER-----RQQLPSGSF 97

QY 201 LFSVYINQLRSDAQYLCQAGDDNSNKKNADLOVLKPEPELYVEDLRGVSFTFCALG 260

Db 98 LIONVVRHRRPDEGVYCEASLDSVGTIVRTAKVLVAGPLRLILSQTESVTF 153

QY 261 PEYANVAKFLCROSSGENDV---VNTLGKRAPEGRILLNPQDKGSFVSFVITGLRK 317

Db 153 --VGDALLRC-EITEGPTISQWNEEDLKVTGDPRLVLP-----SGTLQISRLQ 204

QY 318 EDAGRYLCAHSDQGLQEGSPIQAWOLFVNEESTIPRS-----PTVYKGVAGSSVAVLC 371

Db 205 ADGVYRCLAKNPGSARVNEA---ELRILSEGLRQVFLQRPNSVVAIEGQDAVLEC 261

QY 372 PYNRKESKIKYWCLEGGNQRCPLVDSEGWKAYEGRSLLEPEPNGFTVILNOL 431

Db 262 AVSGYPTPI---VW---MQGDEPPIRTR-----KYSVL-----GGSNLLISN 300

QY 432 TSRDAGFYWCLT---NGDTLWRTTVEIKIEGEPNLKVPGNVAVLGETILKVPCHPCKF 488

Db 301 TDDAGAYTCVATYKKNENTSF--SADLTVMVPPQFLNHPANLYAY--ESMDI--EFECAY 354

QY 489 SSYEKYCKWNNTGCOALPSQDEGSKAFVNCNDSRLVSLTNLVTTRADEGWYKGVKQ 548

Db 355 SGKPSPTVKTNGEVVIPS-----YFQIVDGSNLRILGL---VKSDEGYQC----- 401

QY 549 GHFYGETAAVYAVVERKKAAGSDVS--LAKADAAPDEKVLDSGFREI 594

Db 401 -----IAENEAGNIQTYAQLIPDPAVPSSSILPSAPRDV 435

RESULT 13

B41286

vascular cell adhesion molecule 1, short form precursor - human

C:Species: Homo sapiens (man)

C:Date: 03-Apr-1992 #sequence_revision 14-Jul-1994 #text_change 10-Sep-1997

C:Accession: B41288; A33758

R:Cybulsky, M.I.; Fries, J.W.U.; Williams, A.J.; Sultan, P.; Eddy, R.; Byers, M.; Shows,

Proc. Natl. Acad. Sci. U.S.A. 88, 7859-7863, 1991

A:Title: Gene structure, chromosomal location, and basis for alternative mRNA splicing o

A:Reference number: A41288; MUID:91352090

A:Accession: B41288

A:Molecule type: DNA

A:Residues: 1-647 <CYB>

A:Cross-references: GB:I73255

R:Osborn, L.; Hession, C.; Tizard, R.; Vassallo, C.; Luhowsky, S.; Chi-Rosso, G.; Lo
Cell 59, 1203-1211, 1989
A:Title: Direct expression cloning of vascular cell adhesion molecule 1, a cytokine-1
A:Reference number: A33758; MUID:90090619
A:Accession: A33758
A:Molecule type: mRNA
A:Residues: 1-647 <OSB>
A:Cross-references: GB:M30257; NID:g179885; PID:g179886
C:Keywords: alternative splicing; cell adhesion; glycoprotein; transmembrane protein
F:1-24/Domain: signal sequence #status predicted <SIG>
F:25-647/Product: vascular cell adhesion molecule 1, short form #status predicted <MA
F:25-606/Domain: extracellular #status predicted <EXT>
F:607-628/Domain: transmembrane #status predicted <TMM>
F:629-647/Domain: intracellular #status predicted <INT>
F:273,325,371,439,469/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 5.3%; Score 169.5; DB 2; Length 647;

Best Local Similarity 21.3%; Pred. No. 0.00015;

Matches 138; Conservative 91; Mismatches 225; Indels 195; Gaps 36;

QY 33 GNSVSICTCYPPTSVNRHTRKYWCROGARGCIT-LISSEGVVSSKYAGRANLTNFPENG 91

Db 40 GDSVSLTC-----STTCESPFPSWRTQIDSPLNK--VTNEGTS 78

QY 92 TFVVMIAQLSQDDSGRYKCGIGINSRGLSFDVLSVSGPGLLNDTKVY---TVDLGRTV 148

Db 79 TLTMPVSGFNEHS--YLCTATCESRKLEKIQVEIYFPP---KDPEIHLSPLEAGKPI 133

QY 149 TINC-----PF-----KTENAK-----RSLYK---QIGLYPVL----- 176

Db 134 TVKCSVADYFPDLRLDILLKGDHLMKSQEFLEADRAKSLETKSLEVTFTPIEDIGKVL 193

QY 176 -----VIDSSGVYNPNY-----GRIRLDIOGTG--- 200

Db 194 VCRAKLHIDENDSVPTVRQAVKELQVISPNTVISVNPSTKLQEGSVVTMCSSEGLPA 253

QY 200 -QLFES-----VVINQLRSDAQYLCQAGDDNSNKKNADL--QVLKP 240

Db 254 PEIFWSKKLDNGNLQHLSGNATLTIANRMEDSGIYVCEGVNLGKNRKEVELIYQAFPR 313

QY 241 EPELVYEDLRG-----SVTFHCALGPVAVNAKFLCROSSGENDVVTNLTGKRAPE 294

Db 314 DPEI---EMSGLYNGSVTVSCV-PSVYFPLDRLEILLKE-----TILENIEFLE 362

QY 295 GRILLNPQDKGFSWVITGLRKEDAGRYL--CGA--HSDGQLQEGSPIQAWQ--LEVNES 350

Db 363 DTDKMSLENKSLWTFTPT---IEDTGKALVCQAKLHDDMEFEFKQSTQTLVNV--- 417

QY 351 TIPSPPTVVKVAGSSVAVLCPYNRKESKIKYKWLCEGAONGRCPLLVDSGEGVKAQYE 410

Db 417 VAPRDTTVL--VSPSSIL-----EEGSSVMTCLSQGFP---APKIL---WSRLPN 460

QY 411 GRLSLLEPGNGTFTVILNQLTSRDAGFYWCL--TNGDTLWRTTVEIKIEGEPNLKVPGN 469

Db 461 GELOPLSE--NATLTLSTKM--EDSGVYLCGEGINQAGRSRKEVELIIQVTPKDIK----- 513

QY 470 VTAVLGETLKV--PCHFPEKFSSEYKQYWC---KWNNTCCQALPS-----QDEG 512

Db 513 LTAPSPESVKESDVIISCTCGNVPETWILKKAETGTVLKSIDGAYTIRKAOLKQKADG 572

QY 513 PSKAFVNCNDSRLVSLTNLVTTRADEGWYKGVKQGHFYGETAAVYVA 561

Db 573 VYECESKNKVGSQLRSLTLDVQGRN-----NKDYFSPPELLVLYFA 613

RESULT 14

S50065

sialoadhesin - mouse

C:Species: Mus musculus (house mouse)

C:Date: 07-May-1995 #sequence_revision 21-Jul-1995 #text_change 10-Sep-1997

C:Accession: S50065

R:Crocker, P.R.; Mucklow, S.; Bouckson, V.; McWilliam, A.; Willis, A.C.; Gordon, S.;

A:Reference number: JH0371; MUID:91086838
A:Accession: JH0371
A:Molecule type: mRNA
A:Residues: 1-847 <WIL>
A:Cross-references: GB:593350; NID:g36090; PID:g36091
A:Experimental source: B lymphocyte
A:Note: the authors translated the codon AAT for residue 358 as Met
C:Comment: This protein mediates cell adhesion among human B cells.
C:Genetics:
A:Gene: GDB:C222
A:Cross-references: GDB:127545; OMIM:107266
A:Map position: 19q13.1-19q13.1
C:Superfamily: immunoglobulin homology
C:Keywords: alternative splicing; B-cell; cell adhesion; dimer; glycoprotein; phospho
F:1-19/Domain: signal sequence status predicted <SIG>
F:20-847/Product: B lymphocyte cell adhesion protein #status predicted <MAT>
F:346-398/Domain: immunoglobulin homology <IMM1>
F:609-661/Domain: immunoglobulin homology <IMM2>
F:688-706/Domain: transmembrane #status predicted <TRA>
F:67.101.112.135.164.231.363.445.448.479.574.634/Binding site: carbohydrate (Asn) (co
F:764.789/Binding site: phosphate (Thr) (covalent) #status predicted

A: Gene: GDB:CDL22
A: Cross-references: GDB:127545; OMIM:107266
A: Map position: 19q13.1-19q13.1
C: Superfamily: immunoglobulin homology
C: Keywords: alternative splicing; B-cell; cell adhesion; dimer; glycoprotein; phospho
F: 1-19/Domain: signal sequence status predicted <SIG>
F: 20-847/Product: B lymphocyte cell adhesion protein #status predicted <MAT>
F: 346-398/Domain: immunoglobulin homology <IMM1>
F: 609-661/Domain: immunoglobulin homology <IMM2>
F: 688-706/Domain: transmembrane #status predicted <TRA>
F: 67,101,112,135,164,231,363,445,448,479,574,634/Binding site: carbohydrate (Asn) (CO)
F: 764,789/Binding site: phosphate (Thr) (covalent) #status predicted

	Query Match	5.2%	Score 168.5;	DB 2;	Length 847;
	Best Local Similarity	22.6%;	Pred. No. 0.00025;		
	Matches 146;	Conservative 75;	Mismatches 222;	Indels 203;	Gaps 42;
Qy	25	PEEVNSVEGNSVITCYPPTSVNRHTRKYWCQARGGCITLISSEGYVSSKVGAGRANL	84		
Db	148	PPEIQ--ESQEVTITCLLNFSCGYGYPQLQWLLEGVPMQAAVTSTLTIKSVFT--RSEL	204		
Qy	85	TNFPF---NGTEVWNTAQLSQDSSGRYKCGLINSRGLSPD--VSLEYSQCGLLNDKVKY	140		
Db	205	KFSPOWSHGKIVT--CQL-QDADGKF-----LSNDTVQLNVKHTPKL--EKVY	249		
Qy	141	TVDL----GRVTINCPFKTEAOKRKSLYKQIGLPVLVIDSGYVNPYTGRIRLDIQ	196		
Db	250	PSDAIVREGDSVTMTC-----VSSS---NPEVTVSWLK-D	282		
Qy	197	GTG----QLLFSVINQLRLSDAGQYLCOAGDSSNSK-KNADLOV-LKPEPELVY----	247		
Db	283	GTSLKKQNTFTLLNREVTQDKQSGKYCCQVNDVGPGRSEEFLOQVYAPFSTVQILHSP	342		
Qy	248	DLRGS-VTFPHC-----ALGPEVANVAKFLCRQSGGENDCVVYNT	285		
Db	343	AVEGSOVEFLCSLANPLPTNTVYHNGKMQRTTEKVHPIKILPHHAGTYSY-VAENI	401		
Qy	286	L--GKRPAFAFEGRIILNPODKGSFSVITG---LRKEDAGRYLCGAHSDG---OLQEGS	337		
Db	402	LGTGQRGPGAE---LDVQYPPKVTIVIQNPMPIREGDTVTLCNYSNNSNPSTRYEWM	457		
Qy	338	PIQANQ-----LFVN---ESTI-----PRSPTV--VKGV-	363		
Db	458	PHGAEESPLGLVKLTQNGWDNTTACRNCWSCNASPVALNVQYAPRDVRVRKIKPLS	517		
Qy	363	---AGSSVAVLCYPYNRKESKSIKWLCEQAQNGRCPLLYVDSEGWKQAQYEGRLSLLEP	419		
Db	518	EIHSONSVSLQCDFFSSHPKEVQFF--WE--KNGR--LL-----GKESQLN--	558		
Qy	420	NGGTTVILNQLTSDRAGFYWCLTN---GDT---LNRVTVEIKIIEGPNLKY---PGNV	470		
Db	558	-----FDSISPEDAGSYSCWVNSIGQTASKAW----TLEVLVAPRRLRVSMSPGD-	605		
Qy	471	TAVLTGLKVPCHFPCKFESSYEKYCKWNTGCOALPSQDEGPGSPKVFNCNDSRLVSLT	530		
Db	605	QVMECKSATLTCESDANPVPVSHYTFDWN--QSLPHHSQ-----KLR	645		
Qy	531	LNLVTRADEGWYWC----GVKOGHYFETAAYVAVE----ERKAAG	569		
Db	646	LEPVKVOHSGAWYCGTSGVGRSPLSTLTIVYSPTIGRRVAVG	691		

Qy 471 TAVLGETLUKVPCHFCKSSYBKYNCKWNNTGCQALPSQDEGSKAFVNCDSNRLSVLT 530
+ + + : : : : : : : : :
Db 605 QVMECKSATLTCESDANPPVSHTWFDWNN--QSLPHHSQ-----KLR 645
+ + + : : : : : : : : :
Qy 531 LNLVTRADGGTWC---GVKGCFYGETAAVVAVE---ERKAAG 569
+ + + : : : : : : : : :
Db 646 LEPKVKQHSGAYWCQGTNSVGKRSPLSLTLYVYSPTIGRRVAVG 691

Wed Nov 24 08:12:50 1999

us-09-095-385-4.rpr

Page 10

Search completed: November 20, 1999, 21:55:27
Job time: 34 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 1999, 22:31:39 ; Search time 16.2 Seconds
(without alignments)
2309.786 Million cell updates/sec

Title: US-09-095-385-4
Perfect score: 3226
Sequence: 1 MLFVITCLLAVFPAISTKS.....SGFREIENKAIDPRLFAEZ 608

Scoring table: BLOSUM62

Searched: 201082 seqs, 61543640 residues

Database : SPTREMBL_10.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phase.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2105.5	65.3	771	11	O70570 mus musculus
2	380.5	11.8	102	6	Q29244 sus scrofa
3	210.5	6.5	1323	13	Q08476 gallus gall
4	207.5	6.4	5198	5	O76518 caenorhabdi
5	202	6.3	6642	5	O01761 caenorhabdi
6	199	6.2	6632	5	O17362 caenorhabdi
7	197	6.1	7962	4	O10465 homo sapien
8	195.5	6.1	1896	4	O60468 homo sapien
9	195.5	6.1	1571	4	O60469 homo sapien
10	185	5.7	1028	11	Q62682 rattus norv
11	185	5.7	1344	11	Q92214 mus musculus
12	182.5	5.7	858	5	O18466 hirudo medi
13	182	5.6	390	4	O60667 homo sapien
14	181.5	5.6	1100	4	O94779 homo sapien
15	181.5	5.6	1026	4	O94780 homo sapien
16	181	5.6	1028	11	P97528 rattus norv
17	179	5.5	1273	5	O44928 caenorhabdi
18	178.5	5.5	1099	11	P97527 rattus norv
19	172	5.3	739	6	Q28260 canis famil
20	172	5.3	1377	11	P97603 rattus norv
21	171	5.3	1461	4	Q92859 homo sapien
22	171	5.3	1461	4	O00340 homo sapien
23	170	5.3	1427	13	Q91562 xenopus lae
24	169	5.2	1694	11	O62230 mus musculus
25	167.5	5.2	301	4	O95100 homo sapien
26	167.5	5.2	1028	11	O07409 mus musculus
27	165.5	5.1	662	4	O60926 homo sapien
28	164	5.1	1021	4	Q93033 homo sapien
29	162	5.0	1493	11	P97798 mus musculus

30	160	5.0	739	11	Q63669
31	158.5	4.9	1264	5	P91767
32	156.5	4.9	276	4	O95944
33	156	4.8	4162	13	Q98918
34	154	4.8	1395	5	O44924
35	153.5	4.8	1026	11	Q62845
36	151.5	4.7	1443	13	Q90610
37	151.5	4.7	1005	13	P79921
38	151	4.7	1235	4	O95428
39	148.5	4.6	1239	5	O61541
40	148.5	4.6	1302	5	O61542
41	147.5	4.6	1021	11	O63198
42	147	4.6	1009	13	O93230
43	146.5	4.5	534	5	Q25403
44	144	4.5	538	6	Q28939
45	143	4.4	1535	5	Q23991

ALIGNMENTS

RESULT 1
O70570 PRELIMINARY; PRT; 771 AA.
AC O70570
DT 01-NOV-1998 (TRENBLrel. 08, Created)
DT 01-NOV-1998 (TRENBLrel. 08, Last sequence update)
DT 01-MAY-1999 (TRENBLrel. 10, Last annotation update)
DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR
DE (PLGR) [CONTAINS: SECRETORY COMPONENT].
GN PIGR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/CBAJ (C57 BLACK 6 X CBA); TISSUE=LIVER;
RX MEDLINE; 95138517.
RA PISKURICH J.F., BLANCHARD M.H., YOUNGMAN K.R., FRANCE J.A.,
RA KAETZEL C.S.;
RT "Molecular cloning of the mouse polymeric Ig receptor. Functional
regions of the molecule are conserved among five mammalian species.";
RL J. Immunol. 154:1735-1747(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=129/SVJ;
RA MARTIN M.G., GUTTERREZ E.M., LAN J.T., LI T.W.H., WANG J.;
Gene 201:189-197(1997).
RL [3]
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=129; TISSUE=LIVER;
RA DE GROOT N., VOLLEBRECHT E., LEE S.H., VERBEET M.P., DE BOER H.A.;
Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
BASOLATERAL SURFACE OF EPITHELIAL CELLS.
CC THE COMPLEX IS THEN TRANSPORTED ACROSS THE CELL TO BE SECRETED AT
THE APICAL SURFACE.
CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
EXTRACELLULAR
(KNOWN AS THE SECRETORY COMPONENT) FROM THE TRANSMEMBRANE
SEGMENT.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS 5 V-LIKE DOMAINS.
DR EMBL; U06431; AAA67440.1; -
DR EMBL; U83434; AAC53585.1; -
DR EMBL; U83427; AAC53585.1; JOINED.
DR EMBL; U83428; AAC53585.1; JOINED.
DR EMBL; U83429; AAC53585.1; JOINED.
DR EMBL; U83430; AAC53585.1; JOINED.
DR EMBL; U83431; AAC53585.1; JOINED.
DR EMBL; U83432; AAC53585.1; JOINED.
DR EMBL; U83433; AAC53585.1; JOINED.

Q63669	rattus norv
P91767	manduca sex
O95944	homo sapien
Q98918	gallus gall
O44924	drosophila
Q63845	rattus norv
Q90610	gallus gall
P79921	xenopus lae
O95428	homo sapien
O61541	drosophila
O61542	drosophila
Q63198	rattus norv
O93230	xenopus lae
Q25403	lymaea sta
Q28939	sus scrofa
Q23991	drosophila

```

DR EMBL: Y16524; CAA76272.1; JOINED.
DR EMBL: Y16525; CAA76272.1; JOINED.
DR EMBL: Y16526; CAA76272.1; JOINED.
DR EMBL: Y16527; CAA76272.1; JOINED.
DR EMBL: Y16528; CAA76272.1; JOINED.
DR EMBL: Y16529; CAA76272.1; JOINED.
DR EMBL: Y16530; CAA76272.1; JOINED.
DR EMBL: Y16531; CAA76272.1; JOINED.
DR EMBL: Y16532; CAA76272.1; JOINED.
DR MGD: MGI:23029; PIGR.
DR PFAM: PF00047; ig: 5.
KW Immunoglobulin fold; Repeat: Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 18
FT CHAIN 19 771 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.
FT CHAIN 19 611 SECRETORY COMPONENT.
FT DOMAIN 19 645 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 646 668 POTENTIAL.
FT DOMAIN 669 771 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 33 117 IG-LIKE V-TYPE DOMAIN 1.
FT DOMAIN 145 227 IG-LIKE V-TYPE DOMAIN 2.
FT DOMAIN 250 331 IG-LIKE V-TYPE DOMAIN 3.
FT DOMAIN 363 447 IG-LIKE V-TYPE DOMAIN 4.
FT DOMAIN 477 553 IG-LIKE V-TYPE DOMAIN 5.
FT DISULFID 40 110 POTENTIAL.
FT DISULFID 152 220 POTENTIAL.
FT DISULFID 257 324 POTENTIAL.
FT DISULFID 370 440 POTENTIAL.
FT DISULFID 484 546 POTENTIAL.
FT CARBOHYD 90 90 POTENTIAL.
FT CARBOHYD 147 147 POTENTIAL.
FT CARBOHYD 170 170 POTENTIAL.
FT CARBOHYD 206 206 POTENTIAL.
FT CARBOHYD 420 420 POTENTIAL.
FT CARBOHYD 471 471 POTENTIAL.
FT CONFLICT 159 159 A -> V (IN REF. 1).
FT CONFLICT 396 396 V -> A (IN REF. 1).
FT CONFLICT 620 620 G -> R (IN REF. 1).
FT SEQUENCE 771 AA; 84998 MW; BE9C7E8 CRC32;

Query Match 65.3%; Score 2105.5; DB 11; Length 771;
Best Local Similarity 64.4%; Pred. No. 1.7e-160;
Matches 396; Conservative 82; Mismatches 128; Indels 9; Gaps 5;

QY 1 MLFVLTCLLAVPAISTKPIFGPEVNSVEGNSVITCYPTSVNRHTRKYWCROGA 60
DB 1 MRLVLTLLVTFVSGVSTKPIFGPEVNSVEGNSVITCYPTSVNRHTRKYWCROGA 60
QY 61 RGCITLISSEGYSSKYAGRANLTNPENGTFVNIATQLSODDSGRYKGLGINSRGLS 120
DB 61 SGMCTLLISSNGYLSKEYSGRANLNPENNTFVNIETQLTQDDTGYKGLGTSNRGLS 120
QY 121 FDYSLEVSQPGLLNDTKVYTVDLGRVITNCPFKTENAKRKSLYKQIGLYPVLVDSS 180
DB 121 FDYSLEVSQVPELPSDTHVTKDIGNVTTECPFKENAPSKSLCKTKNQSCELVDST 180
QY 181 GYVNPNTGRIRLDIOGTGOLLSFVSNQLRLSDAGQYLCOAGDDNSNKKNADLQVLP 240
DB 181 EKNPSYIGRAKLFMKGTDLTVFVNIHSLTHNDAGLYICQAGEGPSADKKNDVQLVLP 240
QY 241 EPELVEDRGSTFHCALGPEVANVAKFLCROSSGNCNVDVNTLGKRAPAFEGRILLN 300
DB 241 EPELLYKDLRSSVTFECDLGREVANEAKYLCRMNK-ETCDVIINTLGKRDPDFEGRILIT 299
QY 301 PQDKDSFSVITCLRKEDAGRYLCAHSDGQLOEGSPIQAWOLFVNEESTIPRSPTVVK 360
DB 300 PKDNGRFSVLITGLRKEDAGHYCCGAHSSGLPQEGPIOTWOLFVNEESTIPNRSVVK 359
QY 361 GVAGSSVAVLPCYNRKESKIKYWCWEGAQNGRCPLLYDSEGWYKAYEGRSLSEEPG 420
DB 360 GVTGGVAIACPNPKESSILKYWCWEGNGHCPVLVGTQAOVQOEYEGRLALFDQPG 419
QY 421 NGTFTVILNQLTSRDAGFYWCLNTGLWRTTVEIKIEG--EPNLKV-PGNVTAVLGET 477

```

```

DB 420 NGTFTVILNQLTSDAGFYWCLNTGLWRTTVEIKIEG--EPNLKV-PGNVTAVLGET 479
QY 478 LKVPCHFPCKFSSSEKYWKWNNTGCOALPQSDQSPSKAFYNCDSNLSVLTNLVTPRA 537
DB 480 FTVSCHYPCFKFYVQEKYWKNSKNGCHILPSSHDEGARQSSVSCDQSSOLVMTLNPVSK 539
QY 538 DEGWTWCGVKQGHFYGETAAYVVAVEERKAAGSRDY----SLAKADAAPDEKVLDSGFR 593
DB 540 DEGWTWCGVKQGHFYGETAAYVVAVEERKAAGSRDY----SLAKADAAPDEKVLDSGFR 598
QY 594 IENKAIQDPRLFAEZ 608
DB 599 KENKAIPIPGPFANE 613

RESULT 2
Q29244 PRELIMINARY; PRT; 102 AA.
AC Q29244;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR (PLGR) [CONTAINS: SECRETORY COMPONENT] (FRAGMENT).
GN PIGR.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Mammalia;
OC Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SMALL INTESTINE;
RX MEDLINE; 96327607.
RA WINTEROE A.K., FREDHOLM M., DAVIES W.;
RT "Evaluation and characterization of a porcine small intestine cDNA library: analysis of 839 clones.";
RL Mamm. Genome 7:509-517(1996).
CC -!- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE. DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE TRANSMEMBRANE SEGMENT.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY.
DR EMBL; F14851; CAA23294.1; -.
KW Immunoglobulin fold; Repeat; Transmembrane; Glycoprotein.
FT NON_TER 1 102
FT NON_TER 102 102
FT SEQUENCE 102 AA; 11205 MW; 17A6FA04 CRC32;

Query Match 11.8%; Score 380.5; DB 6; Length 102;
Best Local Similarity 70.9%; Pred. No. 1.5e-23;
Matches 73; Conservative 13; Mismatches 16; Indels 1; Gaps 1;

QY 234 DLQVLKPELVYEDLRGVSFTHCAGPEVANVAKFLCROSSGNCNVDVNTLGKRAPAF 293
DB 1 DLQVLKPELVYEDLRGVSFTHCAGPEVANVAKFLCROSSGNCNVDVNTLGKRAPAF 60
QY 294 EGRILLNQDQXGFSVVTGLRKEDAGRYLCAHSDGQLOEG 336
DB 61 EGRILLTPRE-NSHFSVHTGLRKEDAGHYLCGHHPDGPKEG 102

RESULT 3
Q08476 PRELIMINARY; PRT; 1323 AA.
ID Q08476
AC Q08476;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE CONNECTIN (TITIN) (FRAGMENT).

```

[illegible]

```

Query Match          6.4%; Score 207.5; DB 5; Length 5198;
Best Local Similarity 19.8%; Pred. No. 3.7e-07;
Matches 123; Conservative 86; Mismatches 201; Indels 211; Gaps 28;

QY      26  EEVNSVEGNSVSTC-----YYPPTSVNRHRTKWKCR--QGARGCGITLISSEGVSSKYA 79
      ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
DB      2768  EEVVARVGDTHLTCNAESSVPLSSV-----TWHADESVQGVIT-----SKYA 2812

QY      80  GRANLTNFPENGTVYVNIAGLSODDGGRYKCGIGINSRGLSFD-VSLEVSQGGPGLLNDTK 138
      :| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      2813  ANEK-----TLNVTNIQLDDEGFYVC-TAVNEAGITKKFKFLVITETPYFLDQOK 2861

QY      139  VYVVDLGRVTINCPKFTENAQRKSLYKQIGLYPVLVIDSSGVVNPNTGTRLDITQGT 198
      :| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      2862  LYPIILGKRUTLDC-----SATGTPPTILFMKDGKRU NESDEV--DIIGS 2905

QY      199  GQLFSVYVINQLRLSDAGQYLCQAGDDSNKKNKADQLVLKPEPELYEDLRGSV----- 254
      :||| : : : : : : : : : : : : : : : : : : : : : : :
DB      2906  -----TLIVDNPQKEVEGRYTCIAENKAGRSRDKMMVEVLLP-PKLSKEWINVEVQAGDP 2959

QY      254  -THCAL-----GPEVANVAKFLCRQSSGENCED-----VVVNT 285
      | : : : : : : : : : : : : : : : : : : : : : : :
DB      2960  LTLCEPIEDTSGVHITWSRQFGDQLDMRAQSSDKSKLYIMQATPEDADSYCIAVND 3019

QY      286  LGRKPAFAPEGRIILLNPO-----GKRAPEPEPEPEPEPEPEPEPEPEPEPEPEPEPEPE 303
      | : : : : : : : : : : : : : : : : : : : : : : :
DB      3020  AGGAEAVQVTVVTPPKIFGDSFSTTEIVADTTLLEIPCRTEGIPPEISWFLDGKPILEM 3079

QY      303  -----DKDGSGFSVVITGLRKEDAGRYLCGAHSD-GQLOEGSPIQAWQLFVNEESTIPRSPT 357

DB      3080  PGVITYKQGLSLRIDNKKPNQOEGRYTCVAENKAGRAEQDTYVE-----IGEPRVVMASE 3134

```

```

358  VVKGVAGSSVAVLCP--YNRKESKSIKYWCLEWGAQNGRCPLLVDSEGWK--AQYEGRLS 414
      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
3135 VMRVVEGRQTTIRCEVFGNPE--PVVNW-----LKDGEPTSDLLQGFSTKLS 3179
      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 415  LLEPGNGFTFVLNOLTSRDAGFYWCL--TN--GDTLHRTTVEIKI---IEGEPNLKVP 468
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3180 YLH-----LRETLADGGTGYTCIATNKAGESQTTDVEVLVPPRIEEDERV---- 3226
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 469  NVTAVLGETLKVCPCHFPCKFSSVEKYWKWNNTGCOALPQSDQSPSKAFVNCDSNLS 528
      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 3226 -LOGKEGNTVWVC-----QVTGEPVPYVTKNGRGEIE 3258
      ||| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY 529  -----ITLNLVTRADEGWYWC 544
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 3259 QFNPVLHIRNATRADEGKYSC 3279
      | : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 5
ID01761 PRELIMINARY; PRT; 6642 AA.
AC 001761;
DT 01-JUL-1997 (TrEMBLrel. 04, Created)
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE C. ELEGANS UNC-89 (GB:U33058) (NID:G1160355).
GN UNC-89.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;
OC Rhabditina; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.
[1]
SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RX MEDLINE: 94150718;
RA WILSON R. J., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M.,
RA BONFIELD J., BURTON J., CONNELL M., COPEY T., COOPER J., COULSON A.,
RA CRAXTON M., DEAN S., DU Z., DURBIN R., FAVELLO A., FULTON L.,
RA GARDNER A., GREEN P., HAWKINS T., HILLIER L., JIER M., JOHNSTON L.,
RA JONES M., KERSHAW J., KIRSTEN J., LAISTER N., LATREILLE P.,
RA LIGHTNING J., LLOYD C., MCMURRAY A., MORTIMORE B., O'CALLAGHAN M.,
RA PARSONS J., PERCY C., REKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R.,
RA SHALDON N., SMITH A., SONNHAMMER E., STADEN R., SULSTON J.,
RA THIERRY-MIEG J., THOMAS K., VAUDIN M., VAUGHAN K., WATERSTON R.,
RA WATSON A., WEINSTOCK L., WILKINSON-SPOAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans";
RT Nature 368:32-38(1994).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2;
RC DU Z., LE T., WILSON R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
[3]
RN SEQUENCE FROM N.A.
RP STRAIN=BRISTOL N2;
RC WATERSTON R.;
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR ENBL: AF003131; AAB54132.1;
DR PFAM: PF00041; fn3; 1;
DR PFAM: PF00047; 1g; 37;
DR PFAM: PF00018; SH3; 1;
SQ SEQUENCE 6642 AA; 732821 MW; 63097C09 CRC32;

```

Query Match	6.3%	Score 202;	DB 5;	Length 6642;
Best Local Similarity	21.6%	Pred. No. 1.5e-06;		
Matches 129;	Conservative 67;	Mismatches 230;	Indels 172;	Gaps 24;
QY	88	PENGIFVFNIAQLSODDSGRVKCGIGINSRGLSFDVSL	-----EYSQSGGLINDTK	138
DB	5030	PESGEFSLTIPSSKSDGGAIRVVLG-NDRGEVYSGSVVHVHKSASEPISGANFUSPLK	5088	
QY	139	VVTVDLGRVTATNC-----PFK-----TENAQRKSLYKQ	168	

Db	5089	DTEVEGDMLTLOCTIAGGEFFPEVTEWKGVLQKDDRITHRVALDGTATLIRSAKSD	5149
Qy	169	IGLYPLVLIDSSG-----YVNPNTG-----	190
Db	5149	IGQYRYAKNEAGSATSDCKVTVTGEGQSPKPFVPLKGTGAALPGDKKEFNKVRGLP	5208
Qy	190	-----RRLDIQGTGOLLFSVINLRSLSDAGQYLCQAGDSDNSKNK	232
Db	5209	KPTLQWFLNGIPKFDRIITLDDMADGN--YCLTIRDVREEDFGTLKCIANKENGTDETV	5266
Qy	233	ADLOVLKPEPELYVEDLRSGVTFHCALGPE--VANAKFLCROSSGENCDVVYNTLGR	289
Db	5267	CETQOAGGHDDGSRDDRLPYPRFNPVPLWDRIRPVGDPMFIECHVDANPTAEVWFKGDKK	5326
Qy	290	APAFEGRIILNPQDKSGSVITGLRKEDAGRYLCGAHSD-GOLOEGSPQAQNL-FVN	347
Db	5327	IEHTAHTERN--TVDGACRIKIIIFESDIDGYVMCAVAVNELGQAEQTATYQVEILEHVE	5384
Qy	348	EESTIPRSPV-----VKGVAGSSVAVLCPNRKESKSIKKVWCLWEGAQNGRCPLLY	399
Db	5385	EKKREYAPKINPLEDKTVNG-GOPIRLSCKVDAPRASVWV-----KDG-LPLRA	5435
Qy	400	DSEGWKAQYEGRLSLLEPGNGTFTVILNQLTSDAGFYWCL-TNGDTLWRTTVEIKII	458
Db	5436	DSR--TSIQVE-----EDGTATLAINDSTEEDIAGYRCVATNAHGTINTSCSV---	5482
Qy	459	EGEPNLKPGNVTAVLGE-----TLKVPCHFPCFKFSSEYKWKWN	499
Db	5482	-----NVKPKQEVKKEGEPPFTKGLVDLWADRGDSFTLK-----CAVTGDPPEIKWY	5531
Qy	500	NTGQALPSODEGSPKAFYNCBENSRLVSLTLNLVTRADEGWGCVGKQGHFYGETAA	557
Db	5532	RNG-QLLRN--GPRVIETSPDGS-CSLTWNSTMSDEGIYCEARENAHGAKTQA	5583
RESULT	6		
Q17362		PRELIMINARY:	PRT: 6632 AA.
ID	Q17362		
AC	Q17362		
DT	01-NOV-1996	(TrEMBLrel. 01, Created)	
DT	01-NOV-1996	(TrEMBLrel. 01, Last sequence update)	
DT	01-MAY-1999	(TrEMBLrel. 10, Last annotation update)	
DE	UNC-89.		
GN	UNC-89.		
OS	Caenorhabditis elegans.		
OC	Eukaryota; Metazoa; Nematoda; Secernentea; Rhabditia; Rhabditida;		
OC	Rhabditia; Rhabditoidea; Rhabditidae; Peloderinae; Caenorhabditis.		
RP	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN-BRISTOL N2;		
RC	MEDLINE; 96180278		
RA	BENIAN G.M., TINSLEY T.L., TANG X., BORODOVSKY M.;		
RT	"The Caenorhabditis elegans gene unc-89, required for muscle M-line		
RT	assembly, encodes a giant modular protein composed of Ig and signal		
RT	transduction domains";		
RL	J. Cell Biol. 132:835-848(1996).		
DR	EWBL; U33058; AAB00542.1;		
DR	PFAM; P00041; fn3; 1.		
DR	PFAM; PF00047; Ig; 37.		
DR	PFAM; PF00018; SH3; 1.		
DR	SEQUENCE 6632 AA; 731909 MW; 9F0B6403 CRC32;		

```
Query Match          6.2% Score 199; DB 5; Length 6632;  
Best Local Similarity 21.6%; Pred.No. 2.5e-06;  
Matches 129; Conservative 66; Mismatches 231; Indels 172; Gaps 24;
```

```
y      88 P E N T W I A Q L S O D D G R Y K C G L I N S R G L S F D V S L -----EVSOGPGILLNDTK 138  
       ||| : | : | : | : | : | : | : | : | :  
b     5020 PSGEPSLIPIPSKSGDGAYRVVLG -NDKGVEYGSVVHVKA KSSEPTSCANFLSLPK 5078  
           ||||| : | : | : | : | : | : | : | : | :  
y     139 VTVDLGRVTWTC-----PFK-----TENAKRKSLYQK 168  
         |:| : |:| : | : | : | : | : |
```

```
Db 5079 DTEVEGDMTLTOCTIAGEPEPEVWEKGVVLQKDDRTMRVALDGTATLRIRSPKSD 5138
QY 169 IGLYPLVLVIDSSG-----YVNPNTG-----190
Db 5139 IGQYVSAKNEAGSATSCKVTVTEQEQSPKPFVPLTKGAALPADKKEFNKVRGLP 5198
QY 190 -----RIRLDIOGTQLLFSVVINOLRLSDAGQYLCQAGDDSNKKN 232
Db 5199 NPTQWFLNGIPKLDRLITLDDMADGN--YCLTIRDVREEDFTLCKIAKNNGTDET 5256
QY 233 ADLOVLKPEPELVYEDLRGVTFFCALGPE---VANVAKFLCROSSGENCDVVNTLGKR 289
Db 5257 CEFOQAGHGDGSRDDLURPPFRFVPLWDRRIPYGDPMFIECHVDANPTAEVWFKDGKK 5316
QY 290 APAFEGRIILNPDKDGSFVSVITGLRKEDAGRYLCAHSD--GOLQSGSPQIAQWL--FVN 347
Db 5317 IEHTAHTEIRN--TVDGACRIKIIPFESDIGVYMCVAVNELGQAEQYQYVEILHVE 5374
QY 348 EESTIPRPTV-----YKVAGSSVAVLCPYNRKESKSIKYCLWEGAQNGRCPLLV 399
Db 5375 EKRREYAPKINPLEDKTVNG--GQIRLSCKVDIAIPRASVWY-----KDG-LPLRA 5425
QY 400 DSGVWAQYEGRLSLLEBPGNGFTVILNQLTSRDAQFWCL--TNGDTLWRTTVEIKII 458
Db 5426 DSR--TSIQYE-----EDGTATLAINSTBEDICAYRCVATNAHGTINTSCSV--- 5472
QY 459 EGPENLKVPGNVTAVLGE-----TLKVPCHPCKFSSYKYWCKWN 499
Db 5472 ----NVPKQEVKEGEEPEFTFKGLVDLWADRGDSFTLK-----CAVTGDFPEPKWY 5521
QY 500 NTGQALPSODEGSKAFVNCDENSRLVSLTLNLVTRADEGWYWCYKQGHFYGETAA 557
Db 5522 RNG-QLLRN---GPRTVIETSPDGS--CSLTVNSTMSDEGIYRCEAENHAKRAKTOA 5573

RESULT 7
Q10465 PRELIMINARY; PRT: 7962 AA.
ID Q10465;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-NOV-1998 (Tremblrel. 08, Last annotation update)
DE TITIN, SKELETAL MUSCLE ISOFORM (EC 2.7.1.-) (CONNECTIN) (FRAGMENT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SKELETAL MUSCLE;
RX MEDLINE; 96026330.
RA LABEIT S., KOLMER B.;
RT "Titins: giant proteins in charge of muscle ultrastructure and
RT elasticity";
RL Science 270:293-296(1995).
CC -!- FUNCTION: THIS GIANT MUSCLE PROTEIN MAY BE INVOLVED IN MUSCLE
CC ASSEMBLY AND IN MAINTAINING THE STRUCTURAL INTEGRITY OF
CC SARCOMERES. MAY HAVE PROTEIN KINASE ACTIVITY.
CC -!- TISSUE SPECIFICITY: MUSCLE-SPECIFIC.
CC -!- ALTERNATIVE PRODUCTS: A NUMBER OF FORMS OF THIS PROTEIN ARE
CC PRODUCED BY ALTERNATIVE SPLICING WHICH DIFFER IN TISSUE
CC DISTRIBUTION.
CC DIFFERENT SIZE TRANSCRIPTS MAY ALSO EXIST WITHIN ANY ONE TISSUE.
CC -!- SIMILARITY: TO THE CATALYTIC DOMAINS OF OTHER SERINE/THREONINE
CC KINASES.
CC -!- SIMILARITY: BELONGS TO IMMUNOGLOBULIN SUPERFAMILY.
CC CONTAINS 90 IMMUNOGLOBULIN C2-LIKE DOMAINS.
DR EMBL; X90569; CAA62189.1;
DR PFAM; PF00047; 1g; 56.
KW Muscle protein; Cytoskeleton; Structural protein; Phosphorylation;
KW Serine/threonine-protein kinase; Alternative splicing; Repeat;
KW Immunoglobulin fold.
FT NON_TER 1
DOMAIN 5618 7792 GLU/LYS/PRO/VAL-RICH.
```

```
FT NON_TER 7962 7962
SQ SEQUENCE 7962 AA; 883018 MW; 01C0B7B0 CRC32;

Query Match 6.1%; Score 197; DB 4; Length 7962;
Best Local Similarity 19.9%; Pred. No. 4.8e-06;
Matches 137; Conservative 91; Mismatches 245; Indels 216; Gaps 31;

QY 10 LAVFPAISTKSPIFGPEEVNSVEGNSVITCYYP--PTSVMNRHTRKYWCROGARGCIT 66
Db 1336 LIIPPSFTKL-----KKMDSIKGSFIDLECIVAGSHPSIQ-----WFKD----- 1377
QY 67 LISSEGVSSKYAGRANLTFNPENGTFVYVNIQAOLSDSGRYKCGLSGRSLFDFVSL 126
Db 1377 --DQEISASEKYK---FSFHONTAF--LEISOLECTDSGTYS--ATNKAG----- 1420
QY 127 VSGPGLLNDTKVYVDLGRVTVINGPFTENAKRKSILYKQIGLPLVLDSSGVYVNP 186
Db 1420 -----HNQCSGHLTVKEPPYFVEKPQSD-----VNPV 1447
QY 187 YTGRIRLDIQGTGOLLF-----SVVINOLRLSDAGYLCQA 222
Db 1448 TRVQLKALYGGTAPMTIKFKONKELHGAARSVWDDTSTSLFAAKATDSGTICQL 1507
QY 223 GDDSNKKNADLQV-----LKPEPELVYEDLRG--SVTFHCAL--GPEVANVAKFLCR 272
Db 1508 SNDVGATSKATLVKPEPPQFIKKPSPVLV---LRNGQSTTFECQITGTPKIRVSWYL-- 1563
QY 273 QSSGENCDVVVTLGKRAFAPEGRIILLNQDKDGSFVSVITGLRKEDAGRYLCAHSDGQ 332
Db 1563 --DGNE---ITAIQKHGISF-----IDGLATFOISGARVENSGETYVCEARND- 1606
QY 333 LQEGSPIAQWLQFVNESIPRSPYVVGAGSSVAVLC-----PY-----NRKESKS 380
Db 1606 ---GTASCSIELKVKPEPTFIRELKPVEVVKYSDVELECEVTGTPFEVTLWLNNEIRS 1662
QY 381 IKYWCL-----WEQAQNGRCPLLVDSSEGVAQYEGRLSLLEEPG-----NG 422
Db 1663 SKYTLTDRVSFVFNHITKCDPSDTGEYQCIVSNEGG--SCSCSTRVALKEPPSFIRKIE 1721
QY 423 TETVILLNQLT-----SRDAGFYW-----CLTNGDTLWRT-----TVEIKIIEGPN- 464
Db 1722 TITVLKASSATFSTVAGSPISITWLKDDQILDENNVIFISFVDSVATLQIRSDVNGCHSG 1781
QY 464 -----LKVPGNV---TAVLGETLKVPCHPCKFSSYKYWCKWN 500
Db 1782 RYTQAKNESGVERCYAFLLVQEPQIIVEKAKSDVTEKDPMTLECVAGTPELKVKWLK 1841
QY 501 TGCQALPSODEGSKAFVNCDENSRLVSLTLNLVTRADEGWYWCYKQGHFYGETAAVYV 560
Db 1842 DGKQIVPSR-----YFSMSFENNVAFPRIOSVMKQDSQGYTFKV--ENDFGSSSCDAYL 1893
QY 561 AVEERKAAGSRDVSLSAKADAAPDEKVLDS 589
Db 1894 RVLQDNIPPSFTKLLTKMD-----KVLGS 1917

RESULT 8
O60468 PRELIMINARY; PRT: 1896 AA.
ID O60468;
DT 01-AUG-1998 (Tremblrel. 07, Created)
DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
DT 01-MAY-1999 (Tremblrel. 10, Last annotation update)
DE DOWN SYNDROME CELL ADHESION MOLECULE (FRAGMENT).
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;
RC YAMAKAWA K., HUO Y.-K., HAENDEL M.A., HUBERT R., CHEN X.-N.,
```


RA LYONS G.E., KORENBERG J.R.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF023449; AAC17966.1;
DR PFAM: PF00041; fn3; 6.
DR PFAM: PF00047; fn3; 6.
FT NON_TER 1
SQ SEQUENCE 1896 AA; 209785 MW; 0D6E0CE CRC32;

Query Match 6.1%; Score 195.5; DB 4; Length 1896;
Best Local Similarity 19.3%; Pred. No. 7.6e-07;
Matches 127; Conservative 86; Mismatches 240; Indels 205; Gaps 29;

QY 5 VLTCLLAVFAIPATSKPIFGPEEVNSVEG-----NSVITCYPPPTSNNR--HTRKYWC 56
DB 29 LVPCPAGIPPTLRLWLATGEIYDVGIRHVHPNGTLQIFPPSPSFTLHDNTIYC 88
QY 57 -----ROGARGG-----CITLISSEGYVS----- 76
DB 89 TAENPSGKIRSDVHIKAVLREPYTVRVEDOKTMRGNVAVFKCIIPSSVEAYITVVSWEK 148
QY 76 ---SKYAGRANLTNFPENGTFVNVIAQLSODDSGRYKC-----GLGINSRGLSFDVSL 125
DB 149 DTVSLVSGSRFL--ITSTGALYIKQVQ--NEDGLYNIRCTIRHRYGTGTROSNSARLFVSD 205
QY 126 EVSQGPGLLNDTKVYTVDLGRTVTINCFFKTENNAQRKSLYKQIGLYPVLVIDSSGYVNP 185
DB 206 PANSAPSIIDGFDHRKAMAGORVELPC-----KAL-----GHPEP 240
QY 186 NY-----TGRIRLDIOGTGQLLFSVINQLRLSDAGQYLCQAGDDSNKKNAD 234
DB 241 DYRLKDNMPLELSGRFQKVTG-----LLIENIRPSDSGSYVCEVSNRYGTAKVIGR 293
QY 235 LOVLKP-----EPELVYEDLRGVTFFHCALGPEVANAKFLCROSSGENDCVVYVNTLGKR 289
DB 294 LYKQPLKATISPRKVSQVSLSCS-----VTGTEDQE 330
QY 290 APAFEGRIILNPODKGFSFVITGLR-----KEDAGRYLCGAHSDG----- 332
DB 331 LSWYRNGEILNP-----GKNVRITGINHENLIMDHVYKSDGGAYQCFVRKDKLSAQDYVQ 385
QY 332 -QLOEGSP--IQAWQLFVNEESTIPRSPVTVKGVAGSSVAVLPCYNRKESKSIKYWCLWE 388
DB 386 VLEDDGTPKIISAF-----SEKVVSPAEP-----VSLMCNVKGTPLPTIT--WTL-- 429
QY 389 GAONGRCPLLVDSEGWYKAQYEGRLSLLEPENGTFVILNQLTSRDAGFYWCLTN---G 445
DB 429 -----DDDPILKGGSHRISQMITSEGNVYVYLNISSQVDRDGGYRCTANNSAG 477
QY 446 DTLWRTTVEIKIEGPNLKVPGNVTAVLGETLKVPCHPCKFSSEYKWKCKNNNTGCOA 505
DB 478 VLYQARINVR---GPASIRPMKNITAIAGR---DTYIHCVRVIGYPYISIKWKYKNS--NL 529
QY 506 LPSODEGSKAFVNCNDSRLVSLTLNLVTRADEGHTWCGYKQGHFGYGEAAVYVAVE 563
DB 530 LPFNHR--QVAF-----ENNGLTKLS--DVQKEVDEGEYTCNVLPQLSTQSQSVHVTVK 580

RESULT 9
O60469 PRELIMINARY; PRT; 1571 AA.
AC O60469;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT 01-MAY-1999 (TrEMBLrel. 10, Last annotation update)
DE DOWN SYNDROME CELL ADHESION MOLECULE.
GN DSCAM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=BRAIN;

RA YAMAKAWA K., HUO Y.-K., HAENDEL M.A., HUBERT R., CHEN X.-N.,
RA LYONS G.E., KORENBERG J.R.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF023450; AAC17967.1;
DR PFAM: PF00041; fn3; 6.
DR PFAM: PF00047; fn3; 6.
SQ SEQUENCE 1571 AA; 173803 MW; 5F8C77D1 CRC32;

Query Match 6.1%; Score 195.5; DB 4; Length 1571;
Best Local Similarity 19.3%; Pred. No. 5.7e-07;
Matches 127; Conservative 86; Mismatches 240; Indels 205; Gaps 29;

QY 5 VLTCLLAVFAIPATSKPIFGPEEVNSVEG-----NSVITCYPPPTSNNR--HTRKYWC 56
DB 43 LVPCPAGIPPTLRLWLATGEIYDVGIRHVHPNGTLQIFPPSPSFTLHDNTIYC 102
QY 57 -----ROGARGG-----CITLISSEGYVS----- 76
DB 103 TAENPSGKIRSDVHIKAVLREPYTVRVEDOKTMRGNVAVFKCIIPSSVEAYITVVSWEK 162
QY 76 ---SKYAGRANLTNFPENGTFVNVIAQLSODDSGRYKC-----GLGINSRGLSFDVSL 125
DB 163 DTVSLVSGSRFL--ITSTGALYIKQVQ--NEDGLYNIRCTIRHRYGTGTROSNSARLFVSD 219
QY 126 EVSQGPGLLNDTKVYTVDLGRTVTINCFFKTENNAQRKSLYKQIGLYPVLVIDSSGYVNP 185
DB 220 PANSAPSIIDGFDHRKAMAGORVELPC-----KAL-----GHPEP 254
QY 186 NY-----TGRIRLDIOGTGQLLFSVINQLRLSDAGQYLCQAGDDSNKKNAD 234
DB 255 DYRLKDNMPLELSGRFQKVTG-----LLIENIRPSDSGSYVCEVSNRYGTAKVIGR 307
QY 235 LOVLKP-----EPELVYEDLRGVTFFHCALGPEVANAKFLCROSSGENDCVVYVNTLGKR 289
DB 308 LYKQPLKATISPRKVSQVSLSCS-----VTGTEDQE 344
QY 290 APAFEGRIILNPODKGFSFVITGLR-----KEDAGRYLCGAHSDG----- 332
DB 345 LSWYRNGEILNP-----GKNVRITGINHENLIMDHVYKSDGGAYQCFVRKDKLSAQDYVQ 399
QY 332 -QLOEGSP--IQAWQLFVNEESTIPRSPVTVKGVAGSSVAVLPCYNRKESKSIKYWCLWE 388
DB 400 VLEDDGTPKIISAF-----SEKVVSPAEP-----VSLMCNVKGTPLPTIT--WTL-- 443
QY 389 GAONGRCPLLVDSEGWYKAQYEGRLSLLEPENGTFVILNQLTSRDAGFYWCLTN---G 445
DB 443 -----DDDPILKGGSHRISQMITSEGNVYVYLNISSQVDRDGGYRCTANNSAG 491
QY 446 DTLWRTTVEIKIEGPNLKVPGNVTAVLGETLKVPCHPCKFSSEYKWKCKNNNTGCOA 505
DB 492 VLYQARINVR---GPASIRPMKNITAIAGR---DTYIHCVRVIGYPYISIKWKYKNS--NL 543
QY 506 LPSODEGSKAFVNCNDSRLVSLTLNLVTRADEGHTWCGYKQGHFGYGEAAVYVAVE 563
DB 544 LPFNHR--QVAF-----ENNGLTKLS--DVQKEVDEGEYTCNVLPQLSTQSQSVHVTVK 594

RESULT 10
Q62682 PRELIMINARY; PRT; 1028 AA.
AC Q62682;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-NOV-1998 (TrEMBLrel. 08, Last annotation update)
DE BIG-1 PROTEIN PRECURSOR.
GN BIG-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WISTAR; TISSUE=BRAIN;


```

Db 486 QLKTTIIVTKDQEVVIEC-----KPGSGPKPTISWKKGDRVRENKRIAI 530
Qy 300 NPODKGDSFVSVIITGLRKEDAGRYLCGAHSDGQLQGSGPIQAWQLFVNBEESTIPRSPTVY 359
Db 531 LP---DGSLRL--NASKSDGKYVC---RGENVFGSAEIIASLSVKESPRIELTPKRT 581
Qy 360 KGVAGSVAVLUCPYNRKESKSIK-YWCLWEGAONGRCPLLDVSEGWNKAOYEGRLSJLEE 418
Db 582 ELTVGESIVLNCKAIHSDASLDVIFYMTL-----KGQPIDFEE 618
Qy 419 PGNCTFFVILNQLTSRD-----AGFYWC--LTNGDTLWRTTVEIKIIEGPNLKV 466
Db 619 EG-GHFESIRAAQASADLMIRNILLMHAGRYGCRVQTADV-SDEAEL-LVRGPGG--P 673
Qy 467 PGNV-----TAVLGETLKVCPHFCKFESSY 491
Db 674 PGVIVEETESTATLSWSPAADNHSP--ISSY 704

RESULT 15
O94780 PRELIMINARY; PRT; 1026 AA.
AC AC O94780;
DT 01-MAY-1999 (T-EMBLrel. 10, Created)
DT 01-MAY-1999 (T-EMBLrel. 10, Last sequence update)
DT 01-MAY-1999 (T-EMBLrel. 10, Last annotation update)
DE HNB-2S.
GN HNB-2S.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
OC Eutheria; Primates; Catarrhini; Hominidae; Homo.
[]]
RN RN SEQUENCE FROM N.A.
RC RC STRAIN=CAUCASIAN;
RA KAMEI Y., TSUTSUMI O., TAKETANI Y., WATANABE K.;
RT "cDNA cloning and chromosomal localization of neural adhesion
RT molecule NB-2 in human.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBSJ databases.
DR EMBL; AB013803; BAA36580.1;
SQ SEQUENCE 1026 AA; 112676 MW; 235338211 CRC32;

```

Search completed: November 20, 1999, 23:00:52
Job time: 1753 sec

GenCore version 4.5
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 20, 1999, 21:50:50 ; Search time 10.25 seconds
(without alignments)
1676.792 Million cell updates/sec

Title: US-09-095-385-4

Perfect score: 3226

Sequence: 1 MLLFVLTCLLAVFAISTKS.....SGFREINKAIQDPLFAEZ 608

Scoring table: BLOSUM62

Searched: 77977 seqs, 28268293 residues

Database : SwissProt_37.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	3226	100.0	764	1	PIGR_HUMAN
2	2178	67.5	767	1	PIGR_BOVIN
3	2064.5	64.0	769	1	PIGR_RAT
4	1435.5	44.5	773	1	PIGR_RABIT
5	187.5	5.8	4393	1	PGBM_HUMAN
6	187	5.8	224	1	CM35_HUMAN
7	168.5	5.2	847	1	C22B_HUMAN
8	168.5	5.2	739	1	VCA1_HUMAN
9	164.5	5.1	3707	1	PGBM_MOUSE
10	161	5.0	1239	1	NRG_MOUSE
11	160	5.0	739	1	VCA1_RAT
12	159.5	4.9	1052	1	FGR2_DROME
13	154.5	4.8	837	1	NCM2_MOUSE
14	149.5	4.6	837	1	NCM2_HUMAN
15	147.5	4.6	898	1	FAS2_SCHAM
16	147.5	4.6	808	1	FGR4_MOUSE
17	147	4.6	702	1	CCBM_HUMAN
18	146.5	4.5	1088	1	NCAL_XENLA
19	143	4.4	620	1	SMP_COTJA
20	142	4.4	739	1	VCA1_MOUSE
21	140	4.3	1010	1	CONT_CHICK
22	140	4.3	1447	1	DCC_MOUSE
23	140	4.3	1091	1	NCAL_CHICK
24	139.5	4.3	1020	1	CONT_MOUSE
25	139.5	4.3	2871	1	FBN1_HUMAN
26	139.5	4.3	626	1	MAG_RAT
27	137.5	4.3	1115	1	NCAL_MOUSE
28	137.5	4.3	1092	1	NCAL_XENLA
29	137	4.2	1018	1	CONT_HUMAN
30	137	4.2	879	1	FRP_RAT
31	136.5	4.2	1447	1	DCC_HUMAN
32	136.5	4.2	626	1	MAG_MOUSE
33	135.5	4.2	858	1	NCAL_RAT
34	135	4.2	526	1	BGP1_HUMAN
35	134.5	4.2	2871	1	FBN1_BOVIN
36	134	4.2	1336	1	VGR_RAT
37	133	4.1	725	1	NCAR_MOUSE
38	132	4.1	853	1	NCAL_BOVIN
39	130.5	4.0	1257	1	CAML_HUMAN
40	130.5	4.0	1913	1	KMLS_HUMAN
41	130	4.0	521	1	LAG3_MOUSE
42	130	4.0	1284	1	NRCA_CHICK
43	129.5	4.0	2871	1	FBN1_MOUSE

ALIGNMENTS

```

RESULT 1
PIGR_HUMAN
ID PIGR_HUMAN STANDARD; PRT: 764 AA.
AC P01833;
DT 21-JUL-1986 (REL. 01, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (PLGR) [CONTAINS:
DE SECRETORY COMPONENT].
GN PIGR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92039621.
RA KRAJCI P., GRZESCHIK K.H., GEURTS VAN KESSEL A.H., OLAISEN B.,
RA BRANDTZAEG P.;
RT "The human transmembrane secretory component (poly-Ig receptor):
RT molecular cloning, restriction fragment length polymorphism and
RT chromosomal sublocalization.";
RL HUM. GENET. 87:642-648(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92387236.
RA KRAJCI P., KVALE D., TASKEN K., BRANDTZAEG P.;
RT "Molecular cloning and exon-intron mapping of the gene encoding human
RT transmembrane secretory component (the poly-Ig receptor).";
RL EUR. J. IMMUNOL. 22:2309-2315(1992).
RN [3]
RP SEQUENCE OF 72-764 FROM N.A.
RX MEDLINE: 89149795.
RA KRAJCI P., SOLBERG R., SANDBERG M., OYEN O., JAHNSEN T.,
RA BRANDTZAEG P.;
RT "Molecular cloning of the human transmembrane secretory component
RT (poly-Ig receptor) and its mRNA expression in human tissues.";
RL BIOCHEM. BIOPHYS. RES. COMMUN. 158:783-789(1989).
RN [4]
RP SEQUENCE OF 19-577, DISULFIDE BONDS, AND CARBOHYDRATE-BINDING SITES.
RX MEDLINE: 85128981.
RA EIFFERT H., QUENTIN E., DECKER J., HILLEMEIR S., HUFSCHMIDT M.,
RA KLINGMULLER D., WEBER M.H., HILSCHMANN N.;
RT "The primary structure of human free secretory component and the
RT arrangement of disulfide bonds.";
RL HOPPE-SEYLER'S Z. PHYSIOL. CHEM. 365:1489-1495(1984).
RN [5]
RP SEQUENCE OF 19-577.
RX MEDLINE: 91315750.
RA EIFFERT H., QUENTIN E., WIEDERHOLD M., HILLEMEIR S., DECKER J.,
RA WEBER M., HILSCHMANN N.;
RT "Determination of the molecular structure of the human free secretory
RT component.";
RL BIOL. CHEM. HOPPE-SEYLER 372:119-128(1991).
RN [6]
RP SEQUENCE OF 118-138; 212-230; 232-268; 273-288 AND 578-603.
RX MEDLINE: 97379357.
RA HUGHES G.J., FRUTIGER S., SAVOY L.-A., REASON A.J., MORRIS H.R.,
RA JATON J.-C.;
RT "Human free secretory component is composed of the first 585 amino
RT acid residues of the polymeric immunoglobulin receptor.";
RL FEBS LETT. 410:443-446(1997).
RN [7]
RP FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
RP BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN
RP TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.
RP DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
RP EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE

```

44 128.5 4.0 739 1 PEG1_BOVIN P51866 bos taurus
45 128 4.0 848 1 NCAL_HUMAN P13591 homo sapien

CC TRANSMEMBRANE SEGMENT.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 5 V-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC

DR EMBL: S62403; G238236; -
DR EMBL: S43449; G255098; -
DR EMBL: S43437; G255098; JOINED.
DR EMBL: S43441; G255098; JOINED.
DR EMBL: S43442; G255098; JOINED.
DR EMBL: S43443; G255098; JOINED.
DR EMBL: S43444; G255098; JOINED.
DR EMBL: S43445; G255098; JOINED.
DR EMBL: S43446; G255098; JOINED.
DR EMBL: S43447; G255098; JOINED.
DR EMBL: S43448; G255098; JOINED.
DR EMBL: M24559; G514366; -
DR EMBL: A52091; E307278; -
DR PIR: A02112; ORHUGS.
DR PIR: A32263; A32263.
DR PIR: S13453; S13453.
DR MIM: I73880; -
DR PFAM: PF00047; Ig: 5.
KW IMMUNOGLOBULIN FOLD; REPEAT: TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW POLYMORPHISM.
FT SIGNAL 1 18
FT CHAIN 19 764
FT DOMAIN 19 603
FT TRANSMEM 639 661
FT DOMAIN 662 764
FT DOMAIN 19 132
FT DOMAIN 133 238
FT DOMAIN 239 331
FT DOMAIN 332 463
FT DOMAIN 464 577
FT DISULFID 40 110
FT DISULFID 56 64
FT DISULFID 152 220
FT DISULFID 257 325
FT DISULFID 271 279
FT DISULFID 371 441
FT DISULFID 385 395
FT DISULFID 482 544
FT DISULFID 486 520
FT DISULFID 496 503
FT CARBOHYD 83 83
FT CARBOHYD 90 90
FT CARBOHYD 135 135
FT CARBOHYD 186 186
FT CARBOHYD 421 421
FT CARBOHYD 469 469
FT CARBOHYD 499 499
FT VARIANT 580 580
FT CONFLICT 136 136
FT CONFLICT 158 158
FT CONFLICT 208 208
FT CONFLICT 229 229
FT CONFLICT 234 234
FT CONFLICT 241 241
FT CONFLICT 262 262
FT CONFLICT 280 280
FT CONFLICT 392 392
FT CONFLICT 500 500
FT SEQUENCE 764 AA; 83313 MW; 8FA09656 CRC32;
A -> V.
D -> Q (IN REF. 4 AND 5).
N -> D (IN REF. 4 AND 5).
NO -> DE (IN REF. 4 AND 5).
MISSING (IN REF. 4 AND 5).
D -> N (IN REF. 4 AND 5).
E -> Q (IN REF. 4 AND 5).
E -> Q (IN REF. 4 AND 5).
D -> N (IN REF. 4 AND 5).
N -> D (IN REF. 4 AND 5).
N -> D (IN REF. 4 AND 5).

Query Match 100.0%; Score 3226; DB 1; Length 764;
Best Local Similarity 99.8%; Pred. No. 1.5e-23;
Matches 607; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MLLFVLTCLLAVFAIPAISTKSPFPGPEEYNSVEGNSVITCYPPPTSVNRHTRKYWCROGA 60
DB 1 MLLFVLTCLLAVFAIPAISTKSPFPGPEEYNSVEGNSVITCYPPPTSVNRHTRKYWCROGA 60
QY 61 RGGCITLSSGGYSSKAGANLNFPPENGTFVYVNIQLSQDDSGRYKCGLGINSRGLS 120
DB 61 RGGCITLSSGGYSSKAGANLNFPPENGTFVYVNIQLSQDDSGRYKCGLGINSRGLS 120
QY 121 FDSLEVSQGGPGLLNDTKVTVYDGLGRTVINCFFKTEAQRKSLYKOIGLPLVLIDSS 180
DB 121 FDSLEVSQGGPGLLNDTKVTVYDGLGRTVINCFFKTEAQRKSLYKOIGLPLVLIDSS 180
QY 181 GYVNPNTGRIRLDIQGTGQLLFVSVINQLRSLDAGQYLCQAGDDSNKKNADLQVLKP 240
DB 181 GYVNPNTGRIRLDIQGTGQLLFVSVINQLRSLDAGQYLCQAGDDSNKKNADLQVLKP 240
QY 241 EPELVYEDLRGSVTFHFCALGPEVANVAKFLCRQSGGNCDDVVVNTLGRAPAFEGRIILLN 300
DB 241 EPELVYEDLRGSVTFHFCALGPEVANVAKFLCRQSGGNCDDVVVNTLGRAPAFEGRIILLN 300
QY 301 PQDKDGSFVYITGLRKEDAGRYLCGAHSDGLOQSGSPIQAWOLFVNEESTIPRSPPTVVK 360
DB 301 PQDKDGSFVYITGLRKEDAGRYLCGAHSDGLOQSGSPIQAWOLFVNEESTIPRSPPTVVK 360
QY 361 GVAGSSVAVLCPYNNRKSISIKYWCLEGAONGRCPPLVDSGWWYKAYEGRLSLEEPG 420
DB 361 GVAGSSVAVLCPYNNRKSISIKYWCLEGAONGRCPPLVDSGWWYKAYEGRLSLEEPG 420
QY 421 NGTFVILNQLTSRDAGFYWCLTNGDTLWRTVEIKIIEGEPNLKPCGNVAVLGETLKV 480
DB 421 NGTFVILNQLTSRDAGFYWCLTNGDTLWRTVEIKIIEGEPNLKPCGNVAVLGETLKV 480
QY 481 PCHFPCKFSSEYKWKWNNTGCQALPSQDEGSPSKAFYNCDSNLSVLTNLNLTTRADEG 540
DB 481 PCHFPCKFSSEYKWKWNNTGCQALPSQDEGSPSKAFYNCDSNLSVLTNLNLTTRADEG 540
QY 541 WYWCVGKQGHFYGETAAYVVAEERKAAGSRDVSLSAKADAAPDEKVLDSGFREIENKAIQ 600
DB 541 WYWCVGKQGHFYGETAAYVVAEERKAAGSRDVSLSAKADAAPDEKVLDSGFREIENKAIQ 600
QY 601 DPLFAEZ 608
DB 601 DPLFAEZ 608

RESULT 2
PIGR_BOVIN STANDARD; PRT; 757 AA.
AC P81265;
DT 15-DEC-1998 (REL. 37, CREATED)
DT 15-DEC-1998 (REL. 37, LAST SEQUENCE UPDATE)
DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (PLGR) [CONTAINS:
DE SECRETORY COMPONENT].
GN PIGR.
OS BOS TAURUS (BOVINE).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC ARTIODACTYLA; RUMINANTIA; PECORA; BOVIDAE; BOVINAE; BOS.
[1]
RN SEQUENCE FROM N.A. AND ALTERNATIVE SPLICING.
RC TISSUE=SMALL INTESTINE, AND MAMMARY GLAND;
RX MEDLINE; 95186063.
RA KULSETH M.A., KRAJCI P., MYKLEBOST O., ROGNE S.;
RT "Cloning and characterization of two forms of bovine polymeric
RL immunoglobulin receptor cDNA";
RL DNA CELL BIOL. 14:251-256(1995).
RN [2]

RP SEQUENCE FROM N.A.
 RC TISSUE-MAMMARY GLAND;
 RX MEDLINE: 96069604.
 RA VERBET M.P., VERMEER H., WARWEDAM G.C., DE BOER H.A., LEE S.H.:
 RT "Cloning and characterization of the bovine polymeric immunoglobulin
 RL receptor-encoding cDNA."
 CC GENE 164:329-333(1995).
 CC -!- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
 CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN
 CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.
 CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
 CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE
 CC TRANSMEMBRANE SEGMENT.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.
 CC -!- ALTERNATIVE PRODUCTS: TWO FORMS; A LONG AND A SHORT FORM; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: FOUND IN MAMMARY GLAND, JEJUNUM, LUNG,
 CC KIDNEY AND SMALL INTESTINE.
 CC -!- PTM: IN THE ABSENCE OF DIMERIC IGA, SER-727 IS PHOSPHORYLATED
 CC WHICH ALLOWS PIGR TO FUNCTION NORMALLY.
 CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 CC 5 V-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; L04797; G389280; -
 DR EMBL; X81371; G563341; -
 KW IMMUNOGLOBULIN FOLD; REPEAT; TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
 KW POLYMORPHISM; PHOSPHORYLATION; ALTERNATIVE SPLICING.
 FT SIGNAL 1 18
 FT CHAIN 19 757
 FT DOMAIN 19 599
 FT TRANSMEM 633 653
 FT DOMAIN 654 757
 FT DOMAIN 19 132
 FT DOMAIN 133 238
 FT DOMAIN 239 350
 FT DOMAIN 351 462
 FT DOMAIN 463 575
 FT VARSPIC 129 345
 FT DISULFID 40 110
 FT DISULFID 56 64
 FT DISULFID 152 220
 FT DISULFID 257 324
 FT DISULFID 271 279
 FT DISULFID 370 440
 FT DISULFID 384 394
 FT DISULFID 481 543
 FT DISULFID 485 519
 FT DISULFID 495 502
 FT CARBOHYD 83 83
 FT CARBOHYD 420 420
 FT CARBOHYD 468 468
 FT MOD_RES 727 727
 FT VARIANT 29 29
 FT VARIANT 142 142
 FT VARIANT 404 404
 FT VARIANT 413 413
 FT VARIANT 435 435
 SQ SEQUENCE 757 AA; 82434 MW; 26E40BDB CRC32;

Query Match 67.5%; Score 2178; DB 1; Length 757;
 Best Local Similarity 67.2%; Pred. No. 1.6e-148;
 Matches 405; Conservative 80; Mismatches 114; Indels 4; Gaps 3;

QY 6 LTCLAVFFPAISTKSPIFGPPEVNSVSGNSVSTCYPPPTSVNRHTRKYYWCROGARGCI 65
 DB 6 LACLLAIFPVVSMKSPIFGPPEVTSVEGRSVSIKCYPPPTSVNRHTRKYYWCROGARGCT 65
 QY 66 TLSSSGYSSKYAGRANLTNFPENCTFVYVNIQAQLSDSDSGRYKCGLGINSRGLSDVSL 125
 DB 66 TLSSSGYSSDDVYVGRANLTNFPESGTFVYVDSHSLTHKSDGRYKCGLGISSRGLNFDVSL 125
 QY 126 EVSQGPGGLNDTRKYVYVDLGRVTINCPFTENAQRKSLYKQIGLYPVLVIDSSGVNVP 185
 DB 126 EVSQDPAQASHAHYVYVDLGRVTINCPTRANSEKRSKSLCKTIQDQFOVVDSTGVSN 185
 QY 186 NYTGRRLDIQGTGQLLFVSVINQLRSLDAGQVLCQAGDSDSNKNKADLOVLPKPEPELV 245
 DB 186 SYKDRAHISILGNTLVFVSVINRVKSLDAGMYVCOAGDADAKADKINIDLOVLEPEPELV 245
 QY 246 YEDLRGSVTFHCAIGPEVANVAKFLCROSSGENCDVYVNTLGKRAFAFEGRIILLNPDKD 305
 DB 246 YGDLRSVTFDCSLGPEVANVPAFLCQKNGGACNVVINTLGKKAQDFQGRIVSVPKD-N 304
 QY 306 GSFVSVITGLRKEDAGRYLGAHSDQLQEGSPIQAWQLFVNEESTIPRSPTVYKGVAGS 365
 DB 306 GVFVSHTSLRKEDAGRYVCGAQPGEPODGPVQAWQLFVNEETAIPASPSVYKGVYRG 364
 QY 366 SYAVLCPYNRKESKSYKYLWEGAQNGRCPLLYVDSEGVKWAQYEGRLSLLEPGNGTFT 425
 DB 366 SVTSPYCNPKDANSKYCHWEEAQNRCPLVSEGRGLKEQYEGRLALLTEPGNGTYT 424
 QY 426 VILNQLTSDAGRYWCLTNGDTLWRTTVEIKIEGPNLKVGNVTAVLGTLPKVPCHFP 485
 DB 426 VILNQLTSDAGRYWCLTNGDTLWRTTVEIKIEGPNLKVGNVTAVLGTLPKVPCHFP 484
 QY 486 CKFSYKWKWNNTGCOALPSQDEGSKAFVNCNDSRLSLTLNLVTRADEGWYWG 545
 DB 486 CKFYSEKYKWNKSNRCSALPTQNDGPSQAFVSCDONSQVYVSLNLTDTVTKEDEGWYWG 544
 QY 546 VKOGHYGETAAVYVVERKAAGSDVSLAKADAPDEKVLDSGFGRETKENKAIQDPRLF 605
 DB 546 VKEGPRYGETAAVYVAVESR-VKGSQAGQVK--AAPAGAAIQSRAGEIQNKALDDPSFF 601
 QY 606 AEZ 608
 DB 602 AKE 604
 RESULT 3
 PIGR_RAT
 ID PIGR_RAT STANDARD; PRT; 769 AA.
 AC P15083;
 DT 01-APR-1990 (REL. 14, CREATED)
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
 DT 15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
 DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (PLGR) [CONTAINS:
 DE SECRETORY COMPONENT].
 GN PIGR.
 OS RATTUS NORVEGICUS (RAT).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; RATTUS.
 [1]
 RN SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE: 89378226.
 RA BANTING G., BRAKE B., BRAGHETTA P., LUZIO J.P., STANLEY K.K.;
 RT "Intracellular targeting signals of polymeric immunoglobulin
 RT receptors are highly conserved between species."
 RL FEBS LETT. 254:177-183(1989).
 CC -!- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
 CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN
 CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.
 CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
 CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE
 CC TRANSMEMBRANE SEGMENT.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.


```

CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 5 V-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X15741; G56465;
DR PIR: S05407; S05407;
DR PFAM: PF00047; 19; 5.
DR HSP: P00702; 1JHL.
DR IMMUNOGLOBULIN FOLD; REPEAT: TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
KW IMMUNOGLOBULIN FOLD; REPEAT: TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL.
FT SIGNAL 1 18
FT CHAIN 19 769 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.
FT CHAIN 19 611 SECRETORY COMPONENT.
FT CHAIN 19 643 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 644 666 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 667 769 IG-LIKE V-TYPE DOMAIN 1.
FT DOMAIN 33 117 IG-LIKE V-TYPE DOMAIN 2.
FT DOMAIN 145 227 IG-LIKE V-TYPE DOMAIN 3.
FT DOMAIN 250 331 IG-LIKE V-TYPE DOMAIN 4.
FT DOMAIN 363 447 IG-LIKE V-TYPE DOMAIN 5.
FT DOMAIN 477 553 POTENTIAL.
FT DISULFID 40 110 POTENTIAL.
FT DISULFID 152 220 POTENTIAL.
FT DISULFID 257 324 POTENTIAL.
FT DISULFID 370 440 POTENTIAL.
FT DISULFID 484 546 POTENTIAL.
FT CARBOHYD 90 90 POTENTIAL.
FT CARBOHYD 135 135 POTENTIAL.
FT CARBOHYD 206 206 POTENTIAL.
FT CARBOHYD 471 471 POTENTIAL.
FT SEQUENCE 769 AA; 84798 MW; 9F9D3506 CRC32;
SQ
Query Match 64.0%; Score 2064.5; DB 1; Length 769;
Best Local Similarity 62.6%; Pred. No. 2.2e-140;
Matches 385; Conservative 96; Mismatches 125; Indels 9; Gaps 5;
QY 1 MLFVLTCLLAVTPATSKPIGPEVNSVEGNSVITCYPTSVNRTKRYWQGA 60
DB 1 MRLSLFALLTVFSGVSTQPIFGPDQVSSIEGNSVITCYPTSVNRTKRYWQGA 60
QY 61 RGCITLISSEGVYSSKACRANLTNPFGTFFVNTAQLSDSGRYKGLGINSRGLS 120
DB 61 NGCATLISNGVLSKYSGRASLNPENSTFVINAHLTQEDTSGYKGLGTTNRGLF 120
QY 121 FDVLSVDSQPGLLNDTKVYTVDLGRVTINCPEKFNENAKRSLYKQIGLYPVLVIDSS 180
DB 121 FDVLSVDSQVPEFNDTHVTKDGRVTTECRKEGNAHSKSLCKRGEACEVVDST 180
QY 181 GYVNPYTGRIKDIOGQLLFSVNIQRLSDAGLYLCOAGDSDNSKKNADLVQPKP 240
DB 181 EYVDPYSKDRAILFMKTSRDFVFNISHLIPSDAGLYVCOAGSGPSADKNNADLVQLEP 240
QY 241 EPELVEDLRGVTFFHCALGPVNAVAKFLCROSSGNCNCDVYNTLGKRAPAGEGRILLN 300
DB 241 EPELLYKDRSSVTFECDLGREVANDAKYLCRNK-ETCDVIINTLGKRDPAFEGRILLT 299
QY 301 PQKDGFSFVITGLRKEDAGRYLCAHSDGQLQEGSPIQAWOLFVNEESTIPRSPTVVK 360
DB 300 PRDNGRFSVLIITGLRKEDAGHQCAGHSSGLPQEGHPVQAWOLFVNEESTIENSRYVK 359
QY 361 GVAGSSVAVLPCPNRKESKISKYVWCLWEQAQRCPLVDSEGWAKYQEGRLSLEEPG 420
DB 360 GVTGGSVAIVCPNPKESSILKYVWCHWEADENGRCPLVGTQALVQEGYEGRLALFDQPG 419
QY 421 NGTFTVILNQLTSRDAQFYCLNGDGLTWTVEIKIIEG--EPNKLKVPNGVTVAVLGET 477
DB 421 NGTFTVILNQLTSRDAQFYCLNGDGLTWTVEIKIIEG--EPNKLKVPNGVTVAVLGET 477

```

```

DB 420 SGAYTVILNQLTQDSGFYWCULTDGSRWRTTIELQVAEATKKPDLEVTVPQNATAVIGET 479
QY 478 LKVPCHFPCKFSSEYKWKYKWNNTCQALPSQDEGPKSAFYVNCDSNRLVSLTLNLYTRA 537
DB 480 FTISCHYPCCKFSYQEKYKWKWSDGCHILPSHDEGAROSSVSCDQSSQIVSMTLNPVKKE 539
QY 538 DEGWVWCVKQGHFVGERAAVYVAVVEERKAAGSRDVSIAKADA-----APDEKVLDSGFRRE 593
DB 540 DEGWVWCVKQGHFVGERAAVYVAVVEER--TRGSPHINPTDANARAKDAPEEAEAMSSVRE 598
QY 594 IENKAIQDPRLFAEZ 608
DB 599 DENKANLDRLEFADE 613
RESULT 4
ID PIGR.RABIT STANDARD; PRT: 773 AA.
AC P01832.
DT 21-JUL-1986 (REL. 01, CREATED)
DT 21-JUL-1986 (REL. 01, LAST SEQUENCE UPDATE)
DE POLYMERIC-IMMUNOGLOBULIN RECEPTOR PRECURSOR (PLGR) [CONTAINS:
DE SECRETORY COMPONENT].
GN PIGR.
OS ORYCTOLAGUS CUNICULUS (RABBIT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC LAGOMORPHA; LEPORIDAE; ORYCTOLAGUS.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 84142246.
RA MOSTOV K.E., FRIEDLANDER M., BLOBEL G.;
RT "The receptor for trans epithelial transport of IgA and Igm contains
RT multiple immunoglobulin-like domains.";
RL NATURE 308:37-43(1984).
RN [2]
RP SEQUENCE OF 87-114 AND 410-428.
RX MEDLINE: 88228032.
RA FRUTIGER S., HUGHES G.J., HANLY W.C., JATON J.-C.;
RT "Rabbit secretory components of different allotypes vary in their
RT carbohydrate content and their sites of N-linked glycosylation.";
RL J. BIOL. CHEM. 263:8120-8125(1988).
CC -!- FUNCTION: THIS RECEPTOR BINDS POLYMERIC IGA AND IGM AT THE
CC BASOLATERAL SURFACE OF EPITHELIAL CELLS. THE COMPLEX IS THEN
CC TRANSPORTED ACROSS THE CELL TO BE SECRETED AT THE APICAL SURFACE.
CC DURING THIS PROCESS A CLEAVAGE OCCURS THAT SEPARATE THE
CC EXTRACELLULAR (KNOWN AS THE SECRETORY COMPONENT) FROM THE
CC TRANSMEMBRANE SEGMENT.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. ALSO SECRETED.
CC -!- POLYMORPHISM: THE SEQUENCE SHOWN IS THAT OF ALLOTYPE T62.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 5 V-LIKE DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X00412; G1596;
DR EMBL: K01291; G165106;
DR PIR: A02111; ORREG.
DR PFAM: PF00047; 19; 5.
DR IMMUNOGLOBULIN FOLD; REPEAT: TRANSMEMBRANE; GLYCOPROTEIN; SIGNAL;
KW POLYMORPHISM.
FT SIGNAL 1 18
FT CHAIN 19 773 POLYMERIC-IMMUNOGLOBULIN RECEPTOR.
FT CHAIN 19 615 SECRETORY COMPONENT.
FT DOMAIN 19 647 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 648 670 POTENTIAL.

```

```

FT DOMAIN 671 773 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 30 136 IG-LIKE V-TYPE DOMAIN 1.
FT DOMAIN 137 243 IG-LIKE V-TYPE DOMAIN 2.
FT DOMAIN 244 350 IG-LIKE V-TYPE DOMAIN 3.
FT DOMAIN 351 456 IG-LIKE V-TYPE DOMAIN 4.
FT DOMAIN 457 558 IG-LIKE V-TYPE DOMAIN 5.
FT DISULFID 46 115 POTENTIAL.
FT DISULFID 155 225 POTENTIAL.
FT DISULFID 260 324 POTENTIAL.
FT DISULFID 369 438 POTENTIAL.
FT DISULFID 478 538 POTENTIAL.
FT CARBOHYD 88 88 IN ALLOTYPES T61.
FT CARBOHYD 108 108 IN ALLOTYPES T62 (PARTIAL) AND T63.
FT CARBOHYD 418 418
FT VARIANT 88 88 K -> N (IN ALLOTYPES T61).
FT VARIANT 94 94 D -> E (IN ALLOTYPES T61).
FT VARIANT 101 108 TVDQLTON -> YLNRLSOS (IN ALLOTYPES T61).
FT VARIANT 110 110 S -> T (IN ALLOTYPES T63).
SQ SEQUENCE 773 AA; 83886 MW; 79840D1F CRC32;

Query Match 44.5%; Score 1435.5; DB 1; Length 773;
Best Local Similarity 49.7%; Pred. No. 2.4e-95;
Matches 316; Conservative 87; Mismatches 186; Indels 47; Gaps 18;

QY 1 MLFLVLCCLLAVFAISTK-----SPIFGPEYNSVEGNSVSTCYPTSVNRHTRKY 54
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MALFLVLCCLLAVFAISTK-----SPIFGPEYNSVEGNSVSTCYPTSVNRHTRKY 60
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 55 WCRGARGGITLSSGYYSSKAGANLTFNPENGTFFVYVNIQLSQDSDGRYKCLGI 114
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 61 WCRGARGGITLSSGYYSSKAGANLTFNPENGTFFVYVNIQLSQDSDGRYKCLGI 119
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 115 NSRGLSDVLSVSGPGLNDTVYVLDGRTVTINCPFKTENAQRKSLYKIGLYPV 174
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 120 NSRGLDFGVNVLVSQKPE--PDDVYVYQYSEYTVITCTFYATRLQKSFYKVEDGLV 177
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 LVLDSSG--YVNPNTGRILDIQGTQLLFSVYVNIQLRLSDAGQYLCQAGDDSNKKN 232
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 178 LIIDSSSKAEPYKGRITLIQSTTAKEFTVIKHLQNLNDAGYVYVYVYVYVYVYVYV 237
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 233 ADQLVLPPELVYEDLGRSVTFHCALGPEVYVYVYVYVYVYVYVYVYVYVYVYVYV 292
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 238 VDLRL--TPGLLYCNLGGSVTFECALDSDANAVASL-RQVRGN--VVIDSGGTDPA 292
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 293 FEGRIILNPQDKGSFVITGLRKEADAGRYLCAGHSDGQLEGSPIQAWOLFVNEESTI 352
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 293 FEGRIILNPQDKGSFVITGLRKEADAGRYLCAGHSDGQLEGSPIQAWOLFVNEESTI 350
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 353 PRSPVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 411
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 351 SRSPVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 408
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 412 RLSLLEPGNGTFTVILNQLTSRAGYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 471
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 409 RLALFEPPNGTFTVILNQLTSRAGYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 467
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 472 AVLGELTLKVPCHPCKFSYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 531
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 468 AVOGPEVITCHPCKFSYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 525
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 532 NLVTRADGMYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 583
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 526 DSVSDEDEGMYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYVYV 581
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 584 EKVLDS-----GFREIENK-----ATQDPLFAEZ 608
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 582 EERAKARCPVRRQYPLSKRLTSCPEPLLAEE 617
D : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 5
PGBM_HUMAN
ID PGBM_HUMAN

STANDARD;

PRT; 4393 AA.

P98160; Q16287;
01-OCT-1996 (REL. 34, CREATED)
01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
15-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE
PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC).
HSPG2.
HOMO SAPIENS (HUMAN).
OS EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RX SEQUENCE FROM N.A.
RX MEDLINE; 92112994.
RA KALLUNKI P., TRYGGVASON K.;
RT "Human basement membrane heparan sulfate proteoglycan core protein: a
RT 467-kD protein containing multiple domains resembling elements of the
RT low density lipoprotein receptor, laminin, neural cell adhesion
RT molecules, and epidermal growth factor.";
RL J. CELL BIOL. 116:559-571(1992).
RN [2]
RP SEQUENCE FROM N.A.
RP TISSUE=SKIN, AND COLON;
RX MEDLINE; 92235084.
RA MURDOCH A.D., DODGE G.R., COHEN I., TUAN R.S., IOZZO R.V.;
RT "Primary structure of the human heparan sulfate proteoglycan from
RT basement membrane (HSPG2/perlecan). A chimeric molecule with multiple
RT domains homologous to the low density lipoprotein receptor, laminin,
RT neural cell adhesion molecules, and epidermal growth factor.";
RL J. BIOL. CHEM. 267:8544-8557(1992).
RN [3]
RP SEQUENCE OF 1018-1472 FROM N.A.
RP TISSUE=COLON;
RX MEDLINE; 91365376.
RA DODGE G.R., KOVALSZKY I., CHU M.L., HASSELL J.R., MCBRIDE O.W.,
RA VI H.F., IOZZO R.V.;
RT "Heparan sulfate proteoglycan of human colon: partial molecular
RT cloning, cellular expression, and mapping of the gene (HSPG2) to the
RT short arm of human chromosome 1.";
RL GENOMICS 10:673-680(1991).
RN [4]
RP SEQUENCE OF 892-1398 FROM N.A.
RP TISSUE=FIBROSARCOMA;
RX MEDLINE; 92120660.
RA KALLUNKI P., EDDY R.L., BYERS M.G., KESTILA M., SHOWS T.B.,
RA TRYGGVASON K.;
RT "Cloning of human heparan sulfate proteoglycan core protein,
RT assignment of the gene (HSPG2) to 1p36.1--p35 and identification of
RT a BamHI restriction fragment length polymorphism.";
RL GENOMICS 11:389-396(1991).
RN [5]
RP SEQUENCE OF 1-21 FROM N.A.
RP MEDLINE; 94052171.
RA COHEN I.R., GRAESSEL S., MURDOCH A.D., IOZZO R.V.;
RT "Structural characterization of the complete human perlecan gene and
RT its promoter.";
RL PROC. NATL. ACAD. SCI. U.S.A. 90:10404-10408(1993).
CC -!- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
CC MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
CC CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
CC PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
CC SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
CC SUBSTRATE FOR CELLS.
CC -!- SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
CC DIMERS OR STELLATE STRUCTURES.
CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES,
CC DIMERS OF STELLATE STRUCTURES.
CC -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
CC AND O-LINKED OLIGOSACCHARIDES.
CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
CC -!- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 22 C2-LIKE DOMAINS.

FT	DNAIN	123	175		PRO-RICH.
FT	DISULFID	43	110		POTENTIAL.
FT	DISULFID	57	65		POTENTIAL.
FT	CARBOHYD	90	90		POTENTIAL.
FT	CARBOHYD	99	99		POTENTIAL.
SQ	SEQUENCE	224 AA;	24830 MW;	FC45DC75 CRC32;	

Query Match 5.8%; Score 187; DB 1; Length 224;
Best Local Similarity 29.7%; Pred. No. 1.2e-06;
Matches 51; Conservative 20; Mismatches 75; Indels 26; Gaps

Qy	356	PTVYKGVAGSSAVLCPYNRKESKSIKYACLWEGAQCRCPLLVDSGHWKAQYEGRSL	415
	: : : :		
Db	28	PMTVAGPVGSISVOCRYEHRHLNKFWC--RPPQILCKDKIVETKG-SAGRNGRVSI	84
	: : : :		
Qy	416	LEBPNGTEFVLNOLTSRDAGFYWCLTNGDTLWRITVEIKIIEGPNLKVPNCNVTA--	473
	: : : :		
Db	85	RDSPANLSFTVTLENLTEDAGTYWC--GVDTPWLDRDPHPIVEVESVFPACITASSP	142
	: : : :		
Qy	473	-----VLGETIKVPCHPFKFFSYEKYCKWNNTGCQALPSQDEGPSKAFVN	519
	: : : :		
Db	143	QSSMGTSPPTKLPVH-----TWPSVTRKSDPEPSPGHSILFSN	181
	: : : :		

```

RESULT 7
C22B_HUMAN STANDARD; PRT; 847 AA.
ID C22B_HUMAN STANDARD; PRT; 847 AA.
AC Q01665; Q92872;
DT 01-JUL-1993 (REL. 26, CREATED)
DT 01-JUL-1993 (REL. 26, LAST SEQUENCE UPDATE)
DT 13-DEC-1998 (REL. 37, LAST ANNOTATION UPDATE)
DE B-CELL RECEPTOR CD22-BETA PRECURSOR (LEU-14) (B-LYMPHOCYTE CELL
DE DE ADHESION MOLECULE) (BL-CAM).
GN CD22.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
OC PRIMATES; CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TONSIL;
RX MEDLINE: 91086838.
RA WILSON G.L., FOX C.H., FAUCI A.S., KEHL J.H.;
RT "CDNA cloning of the B cell membrane protein CD22: a mediator of B-B
RT cell interactions.";
RL J. EXP. MED. 173:137-146(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 93267103.
RA WILSON G.L., NAJFIELD V., KOZLOW E., MENNINGER J., WARD D.,
RA KEHL J.H.;
RT "Genomic structure and chromosomal mapping of the human CD22 gene.";
RL J. IMMUNOL. 150:5013-5024(1993).
RN [3]
RP SEQUENCE FROM N.A.
RA LAMERDIN J.E., MCCREADY P., ADAMSON A.W., BURKHART-SCHULTZ K.,
RA GARCIA E., KYLE A., RAMIREZ M., STILWAGEN S., GARNES J.,
RA DANGANAN L., BRUCE R., QUAN G., MONTGOMERY M., OW D.,
RA KOBAYASHI A., OLSEN A.O., CARRANO A.V.;
RL SUBMITTED (AUG-1996) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -1- FUNCTION: MEDIATES B-CELL B-CELL INTERACTIONS. MAY BE INVOLVED IN
CC THE LOCALIZATION OF B-CELLS IN LYMPHOID TISSUES. BINDS SIYLATED
CC GLYCOPROTEINS; ONE OF WHICH IS CD45.
CC -1- SUBUNIT: HETERODIMER OF AN ALPHA AND A BETA CHAIN.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: B-LYMPHOCTES.
CC -1- PTM: PHOSPHORYLATED BOTH ON THREONINE/SERINE AND TYROSINE.
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS; CD22-ALPHA (AC P20273) AND CD22-
CC BETA (SHOWN HERE); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 6 C2-LIKE AND ONE V-LIKE DOMAINS.
CC -1- SIMILARITY: TO MYELIN-ASSOCIATED GLYCOPROTEIN.
CC -1- DATABASE: NAME=PROW; NOTE=GD guide CD22 entry;

```


RT (VCAM1).";
RL J. BIOL. CHEM. 266:6682-6685(1991).
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 25-226.
RX MEDLINE; 95147978.
RA JONES E.Y., HARLOS K., BOTTOMLEY M.J., ROBINSON R.C., DRISCOLL P.C.,
RA EDWARDS R.M., CLEMENTS J.M., DUDGEON T.J., STUART D.I.;
RT "Crystal structure of an integrin-binding fragment of vascular cell
RT adhesion molecule-1 at 1.8-A resolution.";
RL NATURE 373:539-544(1995).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.
RX MEDLINE; 95296382.
RA WANG J.-H., PEPINSKY R.B., STEHLE T., LIU J.-H., KARPUSAS M.,
RA BROWNING B., OSBORN L.;
RT "The crystal structure of an N-terminal two-domain fragment of
RT vascular cell adhesion molecule 1 (VCAM-1): a cyclic peptide based on
RT the domain 1 C-D loop can inhibit VCAM-1-alpha 4 integrin
RT interaction.";
RL PROC. NATL. ACAD. SCI. U.S.A. 92:5714-5718(1995).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.
RX MEDLINE; 95296382.
RA WANG J.-H., STEHLE T., PEPINSKY R.B., LIU J.-H., KARPUSAS M.,
RA OSBORN L.;
RT "Structure of a functional fragment of VCAM-1 refined at 1.9-A
RT resolution.";
RL ACTA CRYSTALLOGR. D 52:369-379(1996).
CC -|- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
CC INTEGRIN VLA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
CC TRANSDUCTION. THE VCAM1/VLA4 INTERACTION MAY PLAY A
CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE
CC EMIGRATION TO SITES OF INFLAMMATION.
CC -|- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -|- TISSUE SPECIFICITY: EXPRESSED ON INFLAMED VASCULAR ENDOTHELIUM, AS
CC WELL AS ON MACROPHAGE-LIKE AND DENDRITIC CELL TYPES IN BOTH NORMAL
CC AND INFLAMED TISSUE.
CC -|- INDUCTION: BY CYTOKINES (E.G. IL-1, TNF-ALPHA).
CC -|- PGM: SIALOGLYCOPROTEIN.
CC -|- DISEASE: MAY PLAY AN IMPORTANT ROLE IN THE GENESIS OF
CC ARTEROSCLEROSIS AND RHEUMATOID ARTHRITIS.
CC -|- ALTERNATIVE PRODUCTS: AT LEAST TWO FORMS ARE PRODUCED BY
CC ALTERNATIVE SPLICING.
CC -|- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 7 C2-LIKE DOMAINS.
CC -|- DATABASE: NAME=PROW; NOTE=CD guide CD106 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd106.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; X53051; G37649;
DR EMBL; M30257; G179886; ALT_TERM.
DR EMBL; M73255; G340196;
DR EMBL; M60335; G340194;
DR PIR; A33758; A33758.
DR PIR; A39755; A39755.
DR PIR; A41288; A41288.
DR PIR; B41288; B41288.
DR PIR; S11476; S11476.
DR PDB; 1VCA; 15-SEP-95.
DR PDB; 1VSC; 20-JUN-96.
DR MIM; 192225;
DR PFAM; PF00047; Ig; 7.
KW IMMUNOGLOBULIN FOLD; GLYCOPROTEIN; CELL ADHESION; TRANSMEMBRANE;
KW SIGNAL; ALTERNATIVE SPLICING; 3D-STRUCTURE.
FT SIGNAL 1 24 PROBABLE.
FT CHAIN 25 739 VASCULAR CELL ADHESION PROTEIN 1.

FT DOMAIN 25 698 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 699 720 POTENTIAL.
FT DOMAIN 721 739 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 38 91 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 129 198 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 237 287 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 326 379 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 418 496 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 525 575 IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 612 675 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 47 95
FT DISULFID 52 99
FT DISULFID 137 195
FT CARBOHYD 273 273 POTENTIAL.
FT CARBOHYD 365 365 POTENTIAL.
FT CARBOHYD 417 417 POTENTIAL.
FT CARBOHYD 463 463 POTENTIAL.
FT CARBOHYD 531 531 POTENTIAL.
FT CARBOHYD 561 561 POTENTIAL.
FT VARSPLIC 310 402 MISSING (IN SHORT FORM).
SQ SEQUENCE 739 AA; 81276 MW; B206C291 CRC32;

Query Match 5.28; Score 168.5; DB 1; Length 739;
Best Local Similarity 24.18; Pred. No. 0.00012;
Matches 134; Conservative 77; Mismatches 227; Indels 117; Gaps 30;

QY 46 SVNRHRTKYWCROGARGGCIYL-ISSEGYVS-----SKYAGRANLTNPENGTFFVNNIAQ 99
DB 229 SVNPSTKL-----QEGSVTMTCSSEGLPAPEFWSKKLDNGNLHLSGNAT--LTLLIA 280
QY 100 LSQDSDSKYKGLGINSRGLS-FDVSLEVSQGPGLLNDKVTYVDLGRVTVINCPFKTEN 158
DB 281 RMEDSGIYVCE-GVNLIGNRKEVELIVQEKP-----FTVEISGPRIAA----- 326
QY 159 AQKRKSLYKQGLPVLVIDSSGVNPNYGRIRLDTOGTQQLFLSVVINOIRLSADAG-- 217
DB 326 -----QIGDSVMTCSVMGCEPSFSWRTQIDPSLKGKVRSEGTNSTLTLSPSVFE 376
QY 217 ---CYLQAGDSDNSNKKNADLOVLK---PEPELVYEDLRG-SVTFHCALGPEVANVAK 268
DB 377 NEHSYLCVTTCGHHKLEKGIQVELYSFPRDPEIEMSGGLVNGSSVTVSCKY-PSVYPLDR 435
QY 269 FLCRSSGEGNDVVVNTIGKRAPAFEGRIILLNPQDKGFSFWITGLRKEDAGRYL-CGA 327
DB 436 LEILLKGE-----TILENIELEDTDKSLENKSLFMTFPT---IEDTGRALVCOA 485
QY 328 --HSDGQLQEGSPIQAWQ-LFVNNEESTIPRPTVVKGVAGSSVAVLCPYNRKESKSIKW 384
DB 486 KLHIDDMFEFPKQSQSTQTLVNV--VAPRDTTVL--VSPSSIL-----EGSSVNNMT 533
QY 385 CLWGAQNGRCPLLVDESGVKAQYEGRLSLEEPGNGTFTVILNQLTSRDAGFWCL-T 443
DB 534 CLSQGFP--APKIL-----WSRQLPNEGLOPLSE--NATLTLISTKM--EDSGVYLCGI 582
QY 444 NGDTLWRTVEIKIEGPNLKVPGNVTVLGETLKV--PCHFPCFSSYEKYWC---KW 498
DB 583 NQAGRSRKEVELLIQVTPKDIK----LTAPSESVKEGDVTIISCTCGNVETWILLAKK 638
QY 499 NNTCCQALPS-----QDEGSPKAFVNCDENSRLVSLTLNLVTRADEGWYMGV 546
DB 639 AETGDTVLKSIDGAYTIRKAQLKDAGVVECESKNKVGSQLRSLTLDVQGRN----- 691
QY 547 KQHFGYETAAVYVA 561
DB 691 NKDYFPELLVLYEA 705

RESULT 9
PGBM_MOUSE
ID PGBM_MOUSE STANDARD; PRT; 3707 AA.
AC Q05793;
DT 01-NOV-1995 (REL. 32, CREATED)

DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
 DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
 DE BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE
 DE PROTEIN PRECURSOR (HSPG) (PERLECAN) (PLC).
 GN HSPG2.
 OS MUS MUSCULUS (MOUSE).
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA;
 OC RODENTIA; SCIUROGNATHI; MURIDAE; MURINAE; MUS.
 RN [1]
 RP -SEQUENCE FROM N.A.
 RC TISSUE-MELANOMA;
 RX MEDLINE; 92078153.
 RA NOONAN D.M., HORGAN E.A., VALENTE P., CAI S., HORGAN E., SASAKI M.,
 RA YAMADA Y., HASSELL J.R.;
 RT "The complete sequence of perlecan, a basement membrane heparan
 RT sulfate proteoglycan, reveals extensive similarity with laminin A
 RT chain, low density lipoprotein-receptor, and the neural cell adhesion
 RT molecule";
 RL J. BIOL. CHEM. 266:22939-22947(1991).
 RN [2]
 RP SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.
 RX MEDLINE; 89034110.
 RA NOONAN D.M., HORGAN E.A., LEDBETTER S.R., VOGELI G., SASAKI M.,
 RA YAMADA Y., HASSELL J.R.;
 RT "Identification of cDNA clones encoding different domains of the
 RT basement membrane heparan sulfate proteoglycan";
 RL J. BIOL. CHEM. 263:16379-16387(1988).
 CC -!- FUNCTION: THIS PROTEIN IS AN INTEGRAL COMPONENT OF BASEMENT
 CC MEMBRANES. IT IS RESPONSIBLE FOR THE FIXED NEGATIVE ELECTROSTATIC
 CC CHARGE AND IS INVOLVED IN THE CHARGE-SELECTIVE ULTRAFILTRATION
 CC PROPERTIES. IT INTERACTS WITH OTHER BASEMENT MEMBRANE COMPONENTS
 CC SUCH AS LAMININ AND COLLAGEN TYPE IV AND SERVES AS AN ATTACHMENT
 CC SUBSTRATE FOR CELLS.
 CC -!- SUBUNIT: PURIFIED PERLECAN HAS A STRONG TENDENCY TO AGGREGATE IN
 CC DIMERS OR STELLATE STRUCTURES.
 CC -!- SUBCELLULAR LOCATION: EXTRACELLULAR.
 CC -!- TISSUE SPECIFICITY: FOUND IN THE BASEMENT MEMBRANES.
 CC -!- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED
 CC AND O-LINKED OLIGOSACCHARIDES.
 CC -!- SIMILARITY: CONTAINS 4 LDL-RECEPTOR CLASS A DOMAINS.
 CC -!- SIMILARITY: CONTAINS 10.5 LAMININ EGF-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 3 LAMININ DOMAINS IV.
 CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
 CC 15 C2-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 LAMININ G-LIKE DOMAINS.
 CC -!- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; M77174; G200296; -;
 DR EMBL; J04054; G200253; -;
 DR EMBL; J04055; G200301; -;
 DR MGD; MGI:96257; HSPG2.
 DR PROSITE; PS00022; EGF_1; 8.
 DR PROSITE; PS01186; EGF_2; 5.
 DR PROSITE; PS01209; LDLRA_1; 4.
 DR PROSITE; PS01248; LAMININ_TYPE_EGF; 11.
 DR PROSITE; PS00068; LDLRA_2; 4.
 DR PFAM; PF00047; Ig; 14.
 DR PFAM; PF00052; laminin_B; 3.
 DR PFAM; PF00053; laminin_EGF; 8.
 DR PFAM; PF00054; laminin_G; 3.
 DR PFAM; PF00057; ldl_recept_a; 4.
 DR HSPG; P01130; LAJ2.
 DR SIGNAL; BASEMENT MEMBRANE; PROTEOGLYCAN; REPEAT; GLYCOPROTEIN;
 KW HEPARAN SULFATE; LAMININ EGF-LIKE DOMAIN; IMMUNOGLOBULIN FOLD;
 KW EXTRACELLULAR MATRIX; EGF-LIKE DOMAIN.

FT	SIGNAL	1	21	POTENTIAL.
FT CHAIN	22	3707		BASEMENT MEMBRANE-SPECIFIC HEPARAN
FT DOMAIN	22	193		SULFATE PROTEOGLYCAN CORE PROTEIN.
FT DOMAIN	194	403		DOMAIN I (UNIQUE, CONTAINS 3 HS SIDE
FT DOMAIN	404	504		CHAINS).
FT DOMAIN	507	1676		DOMAIN II (4 LDLRA REPEATS).
FT DOMAIN	1677	2980		DOMAIN IIA (1 IGG-REPEAT).
FT DOMAIN	2981	3707		DOMAIN III (SIMILAR TO SHORT ARM OF
FT DOMAIN	194	234		LAMININ A CHAIN).
FT DOMAIN	281	319		DOMAIN IV (SIMILAR TO NEURAL CELL
FT DOMAIN	320	359		ADHESION MOLECULE: 14 IGG REPEATS).
FT DOMAIN	360	403		DOMAIN V (C-TERMINAL G-DOMAIN OF LAMININ
FT DOMAIN	404	504		ALPHA CHAINS AND EGF).
FT DOMAIN	521	530		LDL-RECEPTOR CLASS A 1.
FT DOMAIN	531	730		LDL-RECEPTOR CLASS A 2.
FT DOMAIN	731	763		LDL-RECEPTOR CLASS A 3.
FT DOMAIN	764	813		LDL-RECEPTOR CLASS A 4.
FT DOMAIN	814	871		IG-LIKE C2-TYPE DOMAIN 1.
FT DOMAIN	879	923		LAMININ EGF-LIKE 1 (N-TERMINAL).
FT DOMAIN	924	933		LAMININ EGF-LIKE 2.
FT DOMAIN	934	1125		LAMININ EGF-LIKE 3.
FT DOMAIN	1126	1158		LAMININ EGF-LIKE 4 (INCOMPLETE).
FT DOMAIN	1159	1265		LAMININ EGF-LIKE 5 (N-TERMINAL).
FT DOMAIN	1275	1288		LAMININ EGF-LIKE 6.
FT DOMAIN	1295	1324		LAMININ EGF-LIKE 7.
FT DOMAIN	1325	1334		LAMININ EGF-LIKE 8.
FT DOMAIN	1335	1352		LAMININ EGF-LIKE 9 (N-TERMINAL).
FT DOMAIN	1353	1529		LAMININ EGF-LIKE 10.
FT DOMAIN	1529	1562		LAMININ EGF-LIKE 11.
FT DOMAIN	1563	1612		IG-LIKE C2-TYPE DOMAIN 2.
FT DOMAIN	1613	1670		IG-LIKE C2-TYPE DOMAIN 3.
FT DOMAIN	1671	1771		IG-LIKE C2-TYPE DOMAIN 4.
FT DOMAIN	1772	1865		IG-LIKE C2-TYPE DOMAIN 5.
FT DOMAIN	1866	1954		IG-LIKE C2-TYPE DOMAIN 6.
FT DOMAIN	1955	2049		IG-LIKE C2-TYPE DOMAIN 7.
FT DOMAIN	2050	2148		IG-LIKE C2-TYPE DOMAIN 8.
FT DOMAIN	2149	2243		IG-LIKE C2-TYPE DOMAIN 9.
FT DOMAIN	2245	2343		IG-LIKE C2-TYPE DOMAIN 10.
FT DOMAIN	2344	2436		IG-LIKE C2-TYPE DOMAIN 11.
FT DOMAIN	2437	2532		IG-LIKE C2-TYPE DOMAIN 12.
FT DOMAIN	2533	2619		IG-LIKE C2-TYPE DOMAIN 13.
FT DOMAIN	2620	2720		IG-LIKE C2-TYPE DOMAIN 14.
FT DOMAIN	2721	2809		IG-LIKE C2-TYPE DOMAIN 15.
FT DOMAIN	2810	2895		LAMININ G-LIKE 1 (GLOBULAR DOMAIN V A).
FT DOMAIN	2896	2980		EGF-LIKE 1.
FT DOMAIN	2981	3130		LAMININ G-LIKE 2 (GLOBULAR DOMAIN V B).
FT DOMAIN	3049	3241		HEPARAN SULFATE (POTENTIAL).
FT DOMAIN	3304	3495		HEPARAN SULFATE (POTENTIAL).
FT DOMAIN	3558	3705		HEPARAN SULFATE (POTENTIAL).
FT SITE	65	67		HEPARAN SULFATE (POTENTIAL).
FT SITE	71	73		HEPARAN SULFATE (POTENTIAL).
FT SITE	76	78		HEPARAN SULFATE (POTENTIAL).
FT SITE	3615	3617		HEPARAN SULFATE (POTENTIAL).
FT DISULFID	199	212		BY SIMILARITY.
FT DISULFID	206	225		BY SIMILARITY.
FT DISULFID	219	234		BY SIMILARITY.
FT DISULFID	285	297		BY SIMILARITY.
FT DISULFID	292	310		BY SIMILARITY.
FT DISULFID	304	319		BY SIMILARITY.
FT DISULFID	325	337		BY SIMILARITY.
FT DISULFID	332	350		BY SIMILARITY.
FT DISULFID	344	359		BY SIMILARITY.
FT DISULFID	368	381		BY SIMILARITY.
FT DISULFID	375	394		BY SIMILARITY.
FT DISULFID	388	403		BY SIMILARITY.
FT DISULFID	428	479		BY SIMILARITY.
FT DISULFID	764	773		BY SIMILARITY.
FT DISULFID	766	780		BY SIMILARITY.

```

Db 2722 TVQVTPQLETRNIGASVEFFCAVPNERGTHLWLD---KEG-----GQ 2760
Qy 297 ILLNPQDKQDSFVVIITGLRKEDAGRYLCGAHSD-GQLQEGSPIQAWOLFVNEESTIPRS 355
Db 2761 LPPGHSVDQGLVR--IQNLQNCOGTVYVCAHGPGWQAQATA---QLIVQ-----AL 2807
Qy 356 PTVYKGVAGSVAVLCYPNRKESKIYKWLWEGAQNGRCPLLVDSQGWVKAQVEGRLSL 415
Db 2808 PSVLINVRTSVHVSVVGH-----SVFECLALGDPKPVQT-----WSKV--GGHL-- 2851
Qy 416 LEEPG--NGFTFVILNQLSRDAGFYKC-LTNGDTLWRTTVEIKI-----IEGEPNLKV 466
Db 2851 --RPGIVOSGTI-IRIAHVELADAGQRCRAATNAAGTQSHVLLVQALPOLSTPPEIRV 2907
Qy 467 PGNVTAVLGETLKVPCFHFPCKFSSEYKWKWNNTGQALPSQDEGSPSKAFVNCDSNRL 526
Db 2908 PAGSAAV-----FPCMASGYPYPAITW-----SKVDGDLPPDSRL 2942
Qy 527 VS--LTNLNLVTRADEGWYWCVV--KQGHFYGETAAVYVAVEER 565
Db 2943 ENNMLPLSVRPEDAGTYVCTATNRQKV---KAFAYLOVPER 2982

RESULT 10
NRG_DROME STANDARD; PRT; 1239 AA.
ID NRG_DROME
AC P20241: Q24414.
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)
DE DE NEUROGLIAN PRECURSOR.
GN NRG.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PERYGOTA; DIPTERA; BRACHYCERA; MUSCOMORPHA; EPHYDROIDEA;
OC DROSOPHILIDAE; DROSOPHILA.
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 24-41 AND 737-751.
RX MEDLINE; 90030418.
RA BIEBER A.J., SNOW P.M., HORTSCH M., PATEL N.H., JACOBS J.R.,
RA TRAUQUINA Z.R., SCHILLING J., GOODMAN C.S.;
RT "Drosophila neuroglia: a member of the immunoglobulin superfamily
RT with extensive homology to the vertebrate neural adhesion molecule
RT Li.";
RT CELL 59:447-460(1989).
RN [2]
RP SEQUENCE OF 1182-1239 FROM N.A..
RX MEDLINE; 90262720.
RA HORTSCH M., BIEBER A.J., PATEL N.H., GOODMAN C.S.;
RT "Differential splicing generates a nervous system-specific form of
RT Drosophila neuroglia.";
RT NEURON 4:697-709(1990).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 610-814.
RX MEDLINE; 94213741.
RA HUBER A.H., WANG Y.M., BIEBER A.J., BJORKMAN P.J.;
RT "Crystal structure of tandem type III fibronectin domains from
RT Drosophila neuroglia at 2.0 A.";
RT NEURON 12:717-731(1994).
CC -1- FUNCTION: THIS PROTEIN MAY PLAY A ROLE IN NEURAL AND GLIAL CELL
CC ADHESION IN THE DEVELOPING DROSOPHILA EMBRYO.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: NEURONS AND GLIA IN THE DEVELOPING NERVOUS
CC SYSTEM AND ON SOME OTHER NONNEURONAL TISSUES
CC -1- ALTERNATIVE PRODUCTS: A SHORTER FORM IS PRODUCED BY ALTERNATIVE
CC SPLICING.
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC 6 CS-LIKE DOMAINS. STRONG, TO NEURAL CELL ADHESION MOLECULE L1.
CC -1- SIMILARITY: CONTAINS 5 FIBRONECTIN TYPE III-LIKE DOMAINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation

```


CC C2 LINE DOMAIN.....

CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)
CC or send an email to license@isb-sib.ch).
CC

DR EMBL; M84488; G207643; --
DR PIR; JS06675; JS06675.
DR PFAM; PF00047; lg; 5.
DR HSP; P9320; LVCA.
DR IMMUNOGLOBULIN FOLD; CELL ADHESION; TRANSMEMBRANE;
KW SIGNAL.

	1	24
FT CHAIN	25	739
FT DOMAIN	25	698
FT TRANSMEM	699	720
FT SIGNAL	721	739
FT DOMAIN	38	91
FT DOMAIN	129	198
FT DOMAIN	237	287

PROBABLE.
VASCULAR CELL ADHESION PROTEIN 1.
EXTRACELLULAR (POTENTIAL).
POTENTIAL.
CYTOPLASMIC (POTENTIAL).
IG-LIKE C2-TYPE DOMAIN.
IG-LIKE C2-TYPE DOMAIN.
IG-LIKE C2-TYPE DOMAIN.

FT	DOMAIN	326	379	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	418	496	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	525	575	IG-LIKE C2-TYPE DOMAIN.
FT	DOMAIN	612	675	IG-LIKE C2-TYPE DOMAIN.
FT	CARBOHYD	273	273	POTENTIAL.
FT	CARBOHYD	424	424	POTENTIAL.
FT	CARBOHYD	531	531	POTENTIAL.
FT	CARBOHYD	561	561	POTENTIAL.
FT	CARBOHYD	650	650	POTENTIAL.
SQ	SEQUENCE	739 AA:	81246 MW: 41358127 CRC32:	

Query Match

5.0%; Score 160; DB 1: Length 739;

Best Local Similarity 22.6%; Pred.No. 0.00049;

Matches 130: Conservative 84; Mismatches 208; Indels 154; Gaps:

Qy	32	EGNSVSTICY---YPPTSVARHRTKMYCROGARGGCITLISSEGYSSKIYAGRANITNFP	88
		::	::::
Db	238	EGAAVTWTCASEGLPAPEI-----FWSKK-LDNGVLQLLS-----GNATLT---	278
Qy	89	ENGTFVNNIAQLSODDSGRYKCGLGINSRGL-SFDYSLEVSQGPLLNDTKYIVTDLGRF	147
		::::	::::
Db	278	-----LIAMRMDSGIYVCE-GVNLVGRDKTEVELIVOERP-----FTVDI---	318
Qy	148	VTINCPKFTENAQRKSLYKQIGLYPVLVSDSSGYNNPNTGIRLIDIOCTGOLLESVVI	207
		::::	::::
Db	318	-----SPGSOVAQAGDSVVLTCAAAGCDSPFSESWRTQDSDPLNGEVRDEGAT	365
Qy	208	NQURLSDAG-----QYLCQAGDSNSNKKNADLOVUK-----PEPELYEDLRG-SVTFHC	257
		::::	::::
Db	366	STLTLSPGVEDEHSYLTCTVTCORRKLEKTIQEVISFEPDEPDEISGPLVHGRPVTVC	425
Qy	258	ALGPEYANVAKFLCRQSSGENCDVYNTLCKRAPAEGR-ILLNP--QDKDGSFSVVIIG	314
		::::	::::
Db	426	T---VPNVYPF-----DHLEIEL-----LKGETTLNKLFRBEIGTKSLETKS	465
Qy	315	LR-----KEDAGRYL-CGAHSDGQLQEGSPIQ---AWOLFVNEESTIPRSPTVVKGVAG	364
		::::	::::
Db	466	LEMTFIPTAEDTGKALVCLAKLHSSQMESPKQRQSTQLYVN--VAPKEPTI--WVSP	520
Qy	365	SSVAVLCPPYNRKSKSTIKYWCLEWGAQNGRCPLLVDSQGWKVAQYGRGRLSLLEPONGFI	424
		::::	::::
Db	521	SPVP-----EEGSPVNLTCSSDGFT---PKIL-----WSRQLKNGELQPLSQ--NTTL	564
Qy	425	TVILNQLTSRDAGFYNCL-INGDTLWRTTVEIKIIEG---EPNLKYPGNVAVLGETLKV	480
		::::	::::
Db	565	SFMATKM--EDSGIYVCEGINEAGISKKSVEL-IIQSSKDIOLTVFPSSKVKEGDVTII	621
Qy	481	PCHFPCPKFSSYEKYWC---KWNNTGCOALPS-----QDEGPKSAFYNCNDSR	525
		::::	::::
Db	622	SC--TC--GSVPELWIILKKKATGDMVLKSVNGSVTIRKAQLQDAGVYCESKTEVGSG	677
Qy	526	LVSLTNLVLTADBGWYWCYKQGHFYGETAAVYA	561
Db	678	LRSLTLDVKGEN-----NKDYFSPPELLALYFA	705

STRAIN=CANTON-S;
MEDLINE; 93321617.
RA SHISHIDO E., HIGASHIJIMA S.-I., EMORI Y., SAIGO K.;
RA "Two FGF-receptor homologues of Drosophila: one is expressed in
RT mesodermal primordium in early embryos.";
RL DEVELOPMENT 117:751-761(1993).
RN [2]
RP SEQUENCE OF 1-240 FROM N.A.
RP MEDLINE; 92387542.
RA KLAEMET C., GLAZER L., SHILO B.-Z.;
RA "Breathless, a Drosophila FGF receptor homolog, is essential for
RT migration of tracheal and specific midline glial cells.";
RL GENES DEV. 6:1668-1678(1992).
RN [3]
RP SEQUENCE OF 267-1052 FROM N.A.
RP TISSUE=EMBRYO;
RX MEDLINE; 91184623.
RA GLAZER L., SHILO B.-Z.;
RA "The Drosophila FGF-R homolog is expressed in the embryonic tracheal
RT system and appears to be required for directed tracheal cell
RT extension.";
RL GENES DEV. 5:697-705(1991).
RN [4]
RP SEQUENCE OF 868-923 FROM N.A.
RP MEDLINE; 92008631.
RA SHISHIDO E., EMORI Y., SAIGO K.;
RT "Identification of seven novel protein-tyrosine kinase genes of
RT Drosophila by the polymerase chain reaction.";
RL FEBS LETT. 289:235-238(1991).
CC -!- FUNCTION: MAY BE REQUIRED FOR PATTERNING OF MUSCLE PRECURSOR
CC CELLS. WOULD THIS APPEAR ESSENTIAL FOR GENERATION OF MESODERMAL
CC AND ENDODERMAL LAYERS, INVAGINATIONS OF VARIOUS TYPES OF CELLS,
CC AND CNS FORMATION.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- TISSUE SPECIFICITY: MESODERM.
CC -!- DEVELOPMENTAL STAGE: EMBRYOGENESIS. DFR2 EXPRESSION OCCURS IN
CC ENDODERMAL PRECURSOR CELLS, CNS MIDLINE CELLS AND CERTAIN
CC ECTODERMAL CELLS SUCH AS THOSE OF TRACHEA AND SALIVARY DUCT.
CC -!- SIMILARITY: BELONGS TO THE FIBROBLAST GROWTH FACTOR RECEPTOR
CC FAMILY.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. THE
CC EXTRACELLULAR DOMAIN CONTAINS 5 IG-LIKE DOMAINS.

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; X74031; G397601; -.
DR EMBL; X72830; G313848; -.
DR EMBL; X57745; G7966; -.
DR EMBL; S55971; G236016; -.
DR FLYBASE; FBgn0005592; btl.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_IVR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PFAM; PF00047; ig; 4.
DR PFAM; PF00069; pkinase; 1.
DR HSP; P11362; IFGI.
KW RECEPTOR; GLYCOPROTEIN; TYROSINE-PROTEIN KINASE; ATP-BINDING;
KW TRANSFERASE; PHOSPHORYLATION; TRANSMEMBRANE; IMMUNOGLOBULIN FOLD;
KW SIGNAL.
KW SIGNAL 1 19 POTENTIAL.
FT FT CHAIN 20 1052 FGF RECEPTOR HOMOLOG 2.
FT DOMAIN 20 600 EXTRACELLULAR (POTENTIAL).
FT TRANSEM 601 626 POTENTIAL.
FT DOMAIN 627 1052 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 23 97 IG-LIKE DOMAIN 1.
FT FT DOMAIN 157 224 IG-LIKE DOMAIN 2.
FT FT DOMAIN 255 336 IG-LIKE DOMAIN 3.

Qy 7 TCLAIVFAISTKSPDFGPEVNSVEGNSVITCY-----PPTSVNRHTRKYNCRGAR 61
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 106 TWVLEIYOKLTFRE--VSPQEFK--QGDEAEVYCVRSVSSPAPVSWLYHNEE----- 155
Qy 62 GGCITLISSEGVSSKYAGRANLTNFPNGTFVFNIAQLSDSDGGRYKGLGINSRG--LS 120
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 155 ---VTTISD-----FLAMANNLOILNI---NKSDEGTYRCEGRVAREID 197
Qy 121 F-DVLEYSQGLLNDTKVY--TVDLGRVTINCFFKTENAKRKSLYKQIGLYPLVI 177
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 198 FDIIVIVNPPAISMPQKSNFAERTEEMTFC--RASGPEPAISWFRNG---KLIE 252
Qy 178 DSSGVNPNYTGRIKIDTOGTG-QLLFSVINQLKSLDAGQYLCQAGDDSNKKNADLQ 236
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 253 ENEKVI-----LKGSTELTVRNII--SDGGPYVCRATNKAGEDEKQAFLO 297
Qy 237 V-----LKPEPELYEDLRGSVTFHC--ALGPEVANVAKFLCRSGGNCDDVNTLGR 289
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 298 VVQPHIIOQLKNETTYN--GOVTLVCDAGEPIEI-----TWKRAVDGFTTEGDK 348
Qy 290 APAFEGRIILLPQDKGFSVYITGLRKEDAGRYLCGAHSDGQLQEGSPIOAWOLFVNEE 349
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 349 SP--DGRTEVRGQ--HGSSSLHDKVLKSGSGRYDCEAAS---RIGGHQKSMYLDIEYA 400
Qy 350 STIPSPVTVGVAGSSVAVLCPYNRKESKSIKYNWCLWEGAGNRCPLLVDSGWNVAQY 409
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 401 PKFISNOTIYYISWEGNPNISCDVKSNPPASI--HW-----RRDKLVLPKAKNTNLKTY 453
Qy 410 EGRSLLEPGNGTFTVLNLTSDRDAGYWC 441
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 454 TGRKMILE-----IAPTSDNDFGRYNC 475

RESULT 15
FAS2_SCHAM STANDARD: PRT: 898 AA.
ID FAS2_SCHAM
AC P22648;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-FEB-1994 (REL. 28, LAST SEQUENCE UPDATE)
DT 01-JUN-1994 (REL. 29, LAST ANNOTATION UPDATE)
DE FASCICLIN II PRECURSOR (FAS II).
FAS2
OS SCHISTOCERCA AMERICANA (AMERICAN GRASSHOPPER).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; ORTHOPTERA; Caelifera; ACRIDIDAE; CYRTACANTHRADINAE;
OC SCHISTOCERCA.
[1]
SEQUENCE FROM N.A.
RX MEDLINE; 89043938.
RA HARRELSON A.L., GOODMAN C.S.;
RA SNOW P.M., ZINN K., HARRELSON A.L., MCALLISTER L., SCHILLING J.,
RA BASTIANI M.J., MAKH G., GOODMAN C.S.;
RT "Characterization and cloning of fasciclin I and fasciclin II
glycoproteins in the grasshopper".
RL PROC. NACL. ACAD. SCI. U.S.A. 85:5291-5295(1988).
CC -!- FUNCTION: NEURONAL RECOGNITION MOLECULE. INVOLVED IN A PATHWAY
RECOGNITION FOR AXONS DURING THE DEVELOPMENT OF NERVE FASCICLES.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
5 C2-LIKE DOMAINS.
CC -!- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

CC EMBL; J03789; G160849; ..
DR PIR; A40114; A40114.
DR PIR; B31817; B31817.
DR PFAM; PF00041; fn3; 2.
DR PFAM; PF00047; 19; 5.
KW CELL ADHESION; GLYCOPROTEIN; REPEAT; IMMUNOGLOBULIN FOLD;
KW TRANSMEMBRANE; SIGNAL; NEUROGENESIS.
FT SIGNAL 1 22
BY SIMILARITY.
FT CHAIN 23 898
FASCICLIN II.
FT DOMAIN 23 784
EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 765 782
POTENTIAL.
FT DOMAIN 783 898
CYTOPLASMIC (POTENTIAL).
FT DOMAIN 41 120
IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 149 210
IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 241 307
IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 336 414
IG-LIKE C2-TYPE DOMAIN.
FT DOMAIN 443 516
FIBRONECTIN TYPE-III.
FT DOMAIN 548 611
FIBRONECTIN TYPE-III.
FT DOMAIN 662 730
FIBRONECTIN TYPE-III.
FT DISULFID 48 113
POTENTIAL.
FT DISULFID 156 203
POTENTIAL.
FT DISULFID 248 300
POTENTIAL.
FT DISULFID 343 407
POTENTIAL.
FT DISULFID 450 509
POTENTIAL.
FT CARBOHYD 35 35
POTENTIAL.
FT CARBOHYD 51 51
POTENTIAL.
FT CARBOHYD 149 149
POTENTIAL.
FT CARBOHYD 192 192
POTENTIAL.
FT CARBOHYD 297 297
POTENTIAL.
FT CARBOHYD 328 328
POTENTIAL.
FT CARBOHYD 447 447
POTENTIAL.
FT CARBOHYD 457 457
POTENTIAL.
FT CARBOHYD 580 580
POTENTIAL.
SQ SEQUENCE 898 AA; 99064 MW; 55559D5E CRC32;

Query Match 4.6%; Score 147.5; DB 1; Length 898;
Best Local Similarity 21.3%; Pred. No. 0.005;
Matches 104; Conservative 77; Mismatches 189; Indels 119; Gaps 27;

Qy 128 SQPGLLNDTKVTVDLGRVTINCFFKTENAKRKSLYKQIGLYPLVLDSSG-----YV 193
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 24 SAGLEILPNSNOTKPIGRSMILTCKPNVTN---KNLISQLRW-----TDPGREGVPEFK 74

Qy 184 NPN-YTGRIRLD-IOGTGQLLFSVWINQLRLSDAGQYLCQA-GDDSNKKNADLOVLKP 240
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 75 NPILLKPHIFVDWLPPPGKVLTLMIPELREADTGTYTCSALYSNTKQLSKSVHVRTIMP 134

Qy 241 -----EPELYEDLRGSVTFHCALGPEVANVAKFLCRSGGNCDDVNTLGRAPAFEG 295
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 135 ITWDDAPEEQY-----PTVNETFKIRCVSA--NPPAIVNMR-----DG 172

Qy 296 RILLNQ-----DKDGSFVVITGLRKEDAGRYLCGA--HSDGQLQEGSPIOAWOLFVNEE 349
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 173 HIVETGDRYVEQDG---LTILNVTEMDDGTYTCRAIVATGEMAL-RPIRV-----EV 222

Qy 350 STIPRS---PTVVGKVGAGSSVAVLCPYNRKESKSIKYNWCLWEGAGNRCPLLVDSGWN 405
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 223 HTPQMSGALPPKLEAVEGTDFTAKCAASGKP-----VPRTWI 261

Qy 406 KAQYGRSLLEPGNGTFTVLNQLTSR-----DAGFYWCLT-NGDTLWRTTVEIKIIE 459
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 262 RV--DTARDLTGDRVSADVLGELRIEVRPDEADANYSCTAKNAAGTATATVEVTVV 319

Qy 460 GEPNLKVPNGVTAVLETGLK--VPCH-----FPC-----KFSSYKYNWCKNNTGCCALPS 508
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 320 -RPRIGRFDNISVASKDSEAVLECHATGSPPLVTFRLKSLPNRY-----INGIO--PT 371

Qy 509 QDEGSPKAFVNCNENSLVSLTLNLVTRADEGWYMGCV-----KQGHFYGETAAVYV 560
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

Db 372 EDRTVDGVDSPDGRTRIGKLIISNVLRSDDGLYECIATNKGGEVKKNGHLMVEFKPSFA 431

QY 561 AVEERKAAG 569

Db 432 DTPOKEVWG 440

Search completed: November 20, 1999, 21:55:05
Job time: 255 sec